DOCUMENT RESUME

ED 273 799 CE 045 004

| | |
|---|---|
| AUTHOR | Thomson, Peter |
| TITLE | Student Assessment: A Handbook for TAFE Teachers. |
| INSTITUTION | TAFE National Centre for Research and Development, Payneham (Australia). |
| REPORT NO | ISBN-0-17-006896-X |
| PUB DATE | 86 |
| NOTE | 166p. |
| AVAILABLE FROM | Nelson Wadsworth, 480 La Trobe Street, Melbourne, Victoria 3000, Australia ($14.95 Australian). |
| PUB TYPE | Guides - Non-Classroom Use (055) |
| EDRS PRICE | MF01/PC07 Plus Postage. |
| DESCRIPTORS | Achievement Tests; Adult Education; College Students; Continuing Education; Criterion Referenced Tests; *Educational Assessment; Educational Diagnosis; *Evaluation Methods; Postsecondary Education; *Student Evaluation; Teacher Made Tests; Technical Education; Test Construction; *Test Interpretation; Test Reliability; *Test Theory; *Test Use; Test Validity; Vocational Education |
| IDENTIFIERS | *Australia |

ABSTRACT

This handbook was designed to help Australian technical, adult, and further education (TAFE) teachers, especially those in vocational education areas, to assess and evaluate students. The guide contains seven chapters, each written so it can stand alone. At the beginning of each chapter is an explanation of what is to follow and a statement of the objectives that the chapter aims to cover. Chapters conclude with self-assessment exercises for readers and, where appropriate, suggested answers to the exercises are provided. The guide covers the following subjects: introduction to assessment, objectives of assessment, types of questions, test specifications, item analysis, validity and reliability of assessment instruments, and reporting results. Numerous examples are given throughout the chapters. (KC)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED273799

TAFE

# STUDENT ASSESSMENT:

# A HANDBOOK FOR TAFE TEACHERS

CE045004

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C Todd

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Peter Thomson

# STUDENT ASSESSMENT:

# A HANDBOOK FOR TAFE TEACHERS

Peter Thomson

NELSON WADSWORTH

For Mike, Susie and Sam whose lack of enthusiasm for the subject has helped me keep things in perspective.

The TAFE National Centre for Research and Development was set up as a Company in late 1981 by the State, Territory and Commonwealth Ministers for Education, to undertake TAFE research and development projects of national significance.

Nelson Wadsworth
Thomas Nelson Australia
480 La Trobe Street Melbourne Victoria 3000
for TAFE National Centre for Research and Development Ltd.

First published 1986
Copyright c TAFE National Centre for Research
and Development Ltd. 1986

National Library of Australia
Cataloguing in Publication data:

Thomson, Peter, 1933-
Student assessment.

Includes index.
ISBN 0 17 006896 X.

1. Educational tests and measurements. 2. College students, Rating of. 3. Technical education. I. Title.

374'.0132

Camera-ready copy produced by
TAFE National Centre for Research and Development Ltd.

Printed in Australia by Impact Printing (Vic) Pty Ltd.

## CONTENTS


| | | PAGE |
|---|---|---|
| PREFACE | | vii |
| ACKNOWLEDGEMENTS | | ix |
| CHAPTER I | INTRODUCTION TO ASSESSMENT | 1 |
| CHAPTER II | OBJECTIVES | 23 |
| CHAPTER III | TYPES OF QUESTIONS | 51 |
| CHAPTER IV | TEST SPECIFICATIONS | 93 |
| CHAPTER V | ITEM ANALYSIS | 109 |
| CHAPTER VI | VALIDITY AND RELIABILITY | 125 |
| CHAPTER VII | REPORTING RESULTS | 133 |
| INDEX | | 159 |

# PREFACE

A number of good books about the techniques of student assessment are available on the Australian market. However, with only a few exceptions, these are directed at primary and secondary school teachers and their students.

This book is aimed at a different audience; it is directed at technical and further education teachers, particularly at those teachers working in the vocational areas. It also aims to meet the needs of trainers working in commerce and industry who are increasingly being called up to improve the validity of their assessment procedures.

Readers with a knowledge of other texts on the subject will note that although the ground covered is familiar, new ideas and new emphases are presented. Practical assessment receives special attention as do new approaches to describing student performance.

Each Chapter has been written so that it can stand by itself. Chapters can therefore be read and studied independently of each other.

At the beginning of a Chapter there is an explanation of what is to follow and a statement of the objectives that the Chapter aims to cover. Chapters conclude with self-assessment exercises for the readers and, where appropriate, the suggested answers to the exercises are also provided.

**Peter Thomson**
Adelaide, 1986

**ACKNOWLEDGEMENTS**

The development of this book was initially encouraged by the former Director of the TAFE National Centre, Graham Hermann. Geof Hawke of the NSW Department of TAFE and David Mitchell of Queensland TAFE gave considerable help and advice in the early stages. Both Geof and David provided access to assessment materials developed by their respective departments.

Various drafts were commented upon by colleagues in the different states and the current Director of the Centre, Bill Hall, gave invaluable help in shaping the final product.

The manuscript was prepared by Giulia Reveruzzi with great expertise and Margaret Cominos helped edit the text and organised the references.

# CHAPTER I

## INTRODUCTION TO ASSESSMENT

### WHAT THIS CHAPTER IS ABOUT

This chapter has been written to present a general understanding of what is meant by assessment.

Assessments are very much a part of our daily lives. We are constantly asking ourselves questions involving assessment like 'Will I get to work on time ?' 'Is that something I can afford?' 'Is my tennis game improving?'

The chapter addresses the problem of definition. It provides an educational basis for the term 'assessment' and explains how it differs from 'evaluation'.

It should be remembered that this is an introductory chapter and that many of the ideas or concepts mentioned will be taken up in greater detail later. At this stage it is important to become familiar with these concepts and ideas.

*OBJECTIVES*

*After completing this chapter you should be able to:*

- *explain the meaning of the term 'assessment'*
- *explain the meaning of the term 'evaluation'*
- *distinguish between 'measurement' and 'evaluation'*
- *know that in education, measurements are usually done using tests which may be of a variety of types*
- *distinguish between 'formative' and 'summative' evaluation*
- *distinguish between evaluation studies and research studies*
- *state the difference between 'norm-referenced' and 'criterion-referenced' tests and list the relative strengths and weaknesses of each*
- *appreciate the importance of objectives in any assessment program*

## ASSESSMENT, MEASUREMENT, TESTS AND EVALUATION: THE TERMINOLOGY

Reading about assessment and evaluation can be very confusing because different authors use the terms in different ways. To avoid confusion we will settle for a particular set of definitions and avoid looking at alternative definitions.

First let us consider four related terms—assessment, measurement, tests and evaluation.

### Assessment

Assessment is a most difficult term for educators. For some it means 'to test', for others it means 'to judge' or 'to evaluate'. Outside of education it can mean all three and additional things as well.

But since this chapter is about educational assessment it is possible to narrow our definition somewhat. We can do this by reserving the term assessment for things to do with individual students.

Assessment involves making judgments about students according to some set of rules or standards.

Assessment can involve tests where marks are assigned according to some rules, but it can also be judgmental or based on opinions and feelings.

The first type of assessment, the type based mainly on marks and rules is called OBJECTIVE ASSESSMENT. The second type, based on opinions and feelings is termed SUBJECTIVE ASSESSMENT.

In practice an assessment is rarely purely objective or purely subjective. A true/false question on the surface seems highly objective. However it is important to remember that the decision to ask that question involved the opinion that it was a question that should be asked, and such opinions are subjective.

> *Consider the example of a girl who applies for a typing position with an advertising agency.*
>
> *She is given a typing test and is interviewed by the office manager. Her test results show that she can type at 42 words per minute and makes no more than two errors per page—this is an objective assessment.*
>
> *The office manager looks at her work and sees that it is set out very well. But after the interview she reports that the girl was shy, seemed hesitant when answering questions and dressed in an unsuitable manner.*
>
> *The judgments regarding the setting out, shyness and appearance, if they simply reflect the manager's opinions and values, are all subjective assessments.*

It is, of course, possible to turn subjective assessments into objective assessments. In the above case this would involve writing down checklists of standards for 'setting out', 'levels of shyness' and 'appearance'. The applicants could then be graded using each checklist and producing an objective assessment of the three characteristics.

## Measurement

Measurement is one way of making an assessment. Measurement is the assignment of numbers (marks) or grades according to some set of rules.

> *For example, a mark of 9 out of 10 for an essay is a measurement; a comparable grade for such an essay would probably be an A on a possible range from A to E.*

Measurement may use a variety of instruments such as tests, rating scales and checklists. Each of these can be used to make measurements.

> *For example, it is possible to measure the opinions of students as to the level of maintenance of their teaching areas using a simple 5 point scale as follows:*
>
> *How do you rate the overall MAINTENANCE of the various TEACHING AREAS (classrooms, laboratories, workshops)? Tick (✓) the box that corresponds with your opinion.*
>
> *VERY SATISFACTORY* ☐
>
> *SATISFACTORY* ☐
>
> *BORDERLINE* ☐
>
> *UNSATISFACTORY* ☐
>
> *VERY UNSATISFACTORY* ☐
>
> *Should a large number of students rate maintenance as 'Borderline' to 'Very Unsatisfactory' this is a clear measure of the level of student dissatisfaction with maintenance.*

## Tests

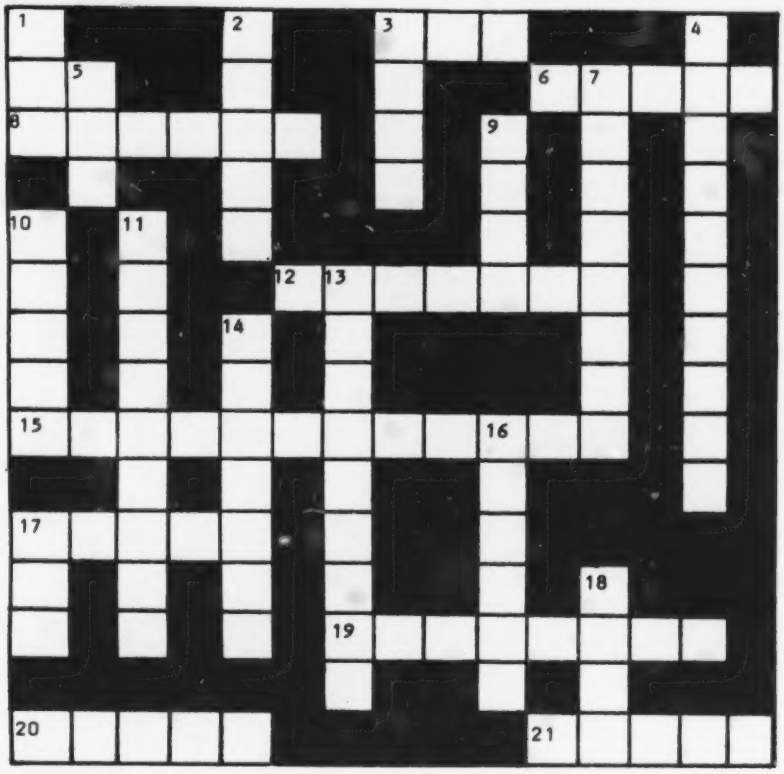
A test is the most commonly used method of making a measurement in education. Tests can take a number of forms, some of which are described below.

### Achievement

Achievement tests measure the amount learned after a period of instruction.

*For example—after a period of instruction on tools used in woodwork, a suitable achievement test to check the student's knowledge of tools might be something like the 'TOOLS CROSSWORD'.*

*TOOLS CROSSWORD*




*CLUES*

*Across*

3. *A tool used when sharpening saws is called a saw* _____

6. *A saw used for the cutting of masonite, plywood etc.*

8. *This chisel must not be hit with a mallet, therefore it is used for fine clean up.*

12. *This tool is made from a flat piece of tool steel 125mm X 75mm and is used for fine clean up of timber surfaces.*

15. *This tool has an adjustable cutter for boring holes up to 75mm.*

17. *Used to turn Auger bits.*

*Across (cont)*

19. *This group of tools reduces the timber surface by scoring or scratching it with a multiple of fine grooves.*

20. *A curved chisel.*

21. *A marking tool used in transferring angles.*

*Down*

1. *This saw is used for cutting with the grain.*

2. *This saw takes its name from a woodworking joint and is used for bench work.*

3. *The name of a cramp used when gluing up frames on the bench.*

4. *A tool used mainly for cutting internal curves where a bow saw cannot be used.*

5. *Used when cutting timber.*

7. *This tool can be either rough cutting or clean cutting, the most common being called a 'Jennings' type.*

9. *This tool is used for quick shaping of curved surfaces.*

10. *Used when marking out.*

11. *Used to remove saw marks etc. from rough sawn timber. The blade is slightly rounded.*

13. *A tool used mainly for cutting scribes in beading and quadrant moulds.*

14. *Two types of this tool are produced, the: a) Tower*
*b) Lancashire*
*and is used mainly for extracting nails or tacks which have become bent in driving.*

16. *A hand tool used for cutting curves in thick timber.*

17. *This tool is called a Pinch .....*

18. *Used in the clean up of curved surfaces.*

*(From Basic Carpentry/Joinery Trade: Basic Module 1.6, SA TAFE)*

**Aptitude**

Aptitude tests are designed to predict a person's future performance.

Aptitude tests involve the solution of problems that are outside the student's direct experience. A test of knowledge **cannot** be an aptitude test.

*Aptitudes with 'number series' provide a suitable example.*

*Although a student might be able to identify the rules for supplying the missing number from the series*

*2, 4, ?, 256 (Answer 16)*

*another 'series' will require the identification of a new rule. So the series*

*2, 4, ?, 5, 4 (Answer 3)*

*is different and can be assumed to be outside the student's direct experience. This classification works reasonably well if the test item involves something the student has not encountered or rarely encountered previously.*

*However, if the solution of 'number series' is a part of the curriculum and the student gets a lot of practice at solving such problems, 'number series' become achievement items rather than aptitude items.*

Many so-called spatial ability tests can be classified as aptitude tests, particularly if spatial ability is something not covered in the curriculum.

*For example:*

*If the shape shown in Figure I is cut out and folded, it forms a solid, the top half of which is shown in Figure II.*

Figure I

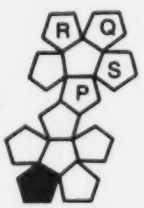


Figure II

*An edge common to the black face occurs with*

*A Q only;*

*B R only;*

*C both Q and R;*

*D neither Q nor R.*

*(Answer C)*

*(ACER MATHS ITEM BANK BOOK 2, 618011).*

**Diagnostic**

Diagnostic tests identify a student's difficulties or patterns of error. A diagnostic test is often used to provide the basis for remedial instruction.

> *For example, in mathematics it is of fundamental importance to be able to recognise the symbols which represent the 'four processes', addition, subtraction, multiplication and division.*
>
> *A series of diagnostic test items which address this issue might begin with:*

| *(Q1) 14 X 7 = ?* | | *(Q2) 15 X 5 = ?* | | *(Q3) 16 ÷ 4 = ?* | |
|---|---|---|---|---|---|
| *A* | *98* | *A* | *3* | *A* | *64* |
| *B* | *21* | *B* | *10* | *B* | *20* |
| *C* | *7* | *C* | *20* | *C* | *12* |
| *D* | *2* | *D* | *75* | *D* | *4* |

> *And an answer pattern such as*
>
> *A, D, D, suggests there are no problems at this level with the symbols X and ÷*
>
> *whereas*
>
> *B, C, D suggests confusion between the symbols X and +*
>
> *and*
>
> *D, A, A suggests confusion between the symbols X and ÷*
>
> *Provided such error patterns are consistent over a sufficiently large number of items the teacher should be able to diagnose particular problems the student is having.*

Tests are further considered under the section on norm-referenced and criterion-referenced assessments pp 10-14.

**Evaluation**

Evaluation is the assessment of the merit or worth of something.

Evaluations are about such things as programs, institutions and policies. Assessments of students are typical components of an educational evaluation. Evaluation can be both 'quantitative' and 'qualitative'.

When an evaluation involves the consideration of numbers (e.g. the percentage of students passing a test) it is called 'quantitative'. When it is based on opinions or values (without reference to numbers, grades etc.) it is called 'qualitative'.

> The **quantitative** information of evaluation consists of elements which have been assigned numbers, grades or ratings of some type. Measurements are quantitative information.

The **qualitative** information used in evaluations usually consists of descriptive statements which aim to elaborate, clarify, and analyse the issues under consideration.

There are two broad functions of evaluation which are categorised as formative and summative.

**Formative evaluation**

Formative evaluation refers to the act of assessing the worth or providing information about something such as a teaching aid, a teaching program, a college etc. with a view to making on-going changes. Implicit in the process of formative evaluation is the requirement that the activities being evaluated can be modified during the course of the evaluation. Formative evaluations are characterised by flexibility and a commitment to making changes (improvements) as soon as these are perceived to be necessary.

**Summative evaluation**

Summative evaluation, on the other hand, assesses the worth or provides information on an activity at some predetermined point in time or on the activity's completion. Summative evaluations have little flexibility. They may be designed with improvement in mind, but any changes aimed at improving things are made at the end of the evaluation.

*Monitoring the use of an interim collection of curriculum materials in a classroom setting before they go for publication in a 'final form' is therefore the task of a formative evaluator.*

*The formative evaluator's job is to ensure that changes are made that improve the materials prior to the production of the 'final form'.*

*After these curriculum materials have been enshrined between hard covers, a judgment about their worth would become the task of a summative evaluator.*

**Evaluation and research**

It is also useful to distinguish research from evaluation. Research studies are those primarily designed to add to the body of knowledge on the subject being researched. Research is about drawing conclusions; it emphasises the value of truth and has high generalisability. Evaluation, on the other hand, is about providing information for particular purposes and is less generalisable.

*An article entitled 'The theory of vocational education' is likely to be about research in TAFE whereas the title 'How electronics is taught in TAFE colleges' suggests an evaluation study.*

**ASSESSMENT IN TAFE**

Assessment can be used for many purposes, but in the context of TAFE, there are two major functions namely:

- to assist in making decisions about the performance of individual students
- to provide information about the teaching and learning processes.

The methods of assessment in a subject should be based upon the answers to two general questions. Firstly, for each set of objectives, what is the most valid, reliable and practicable method of testing? Secondly, how can this method of testing be most appropriately conducted?

In considering these questions the following points need to be taken into account.

- End-of-year examinations are, in many cases, an appropriate means of testing knowledge and understanding of the content of a subject and, to some extent, how this knowledge and understanding can be applied. But such examinations do not allow suitable feedback to the teacher or the student. End-of-year examinations are summative assessments of student performance. Since they come at the end of a period of instruction there is no chance for their results to be used for any on-going improvement of student performance. Of course, the teacher might be able to reflect on the results and do something to improve his or her teaching next time. Unfortunately, this is too late for the group of students who have just been tested.

  Conversely, well-designed class tests can readily provide the required feedback. Results of class tests, if used for improvement in the teaching and learning situation, are formative assessments.

- In many TAFE subjects application of knowledge and understanding is most validly tested by a task that replicates, as far as possible, the situation in the industrial or commercial field for which the subject was designed to prepare the student.

- Generally, TAFE students are more positively motivated by tests relevant to on-the-job performance and by continuous assessment than by tests of knowledge and understanding or by end-of-year examinations. (Continuous assessment involves testing the progress of students in a subject at regular intervals, say for example, on the completion of a unit of work. In some ways the term is an unfortunate one as it is taken [quite wrongly] by some people to mean continual assessment.)

- Continuous assessment is better able to provide feedback on the effectiveness of the educative process to teachers and students than is end-of-year assessment.

After making such considerations, the specific objectives of a subject can be assigned to various methods of assessment such as end-of-year examinations, class tests, assignments, projects, practical tests, orals and aural tests. These various methods of assessment are considered in greater detail in Chapter III.

*Two general rules emerge from the above:*

1. *Continuous assessment is preferable to once only end-of-year examinations because it*
   - *gives a greater motivation to students*
   - *provides regular feedback on performance to both teachers and students.*
2. *Real-life practical tests are preferable to artificial ones.*

**The role of norm-referenced and criterion-referenced assessments**

The following anecdote outlines the difference between these types of assessment:

*Dr Wun after examining a patient who had been in a road accident was asked by the patient's family what was the chance of recovery. Dr Wun shook her head and replied, 'Well, her kidneys have been badly damaged and we haven't been able to stop the internal bleeding. I'm afraid there is little chance for her'.*

*Dissatisfied with this prediction the family sought a second opinion from Dr Too. When asked the same question she smiled reassuringly and said, 'Well she is very low, but compared to the other women in the hospital ward her chances of recovery are quite good'.*

*In making a judgment about the patient's chance of recovery in relation to the criteria of kidney damage and internal bleeding, Dr Wun was making a criterion-referenced assessment.*

*Dr Too's judgment, based on the chances of survival compared with that of other patients in the ward, can be likened to a norm-referenced assessment.*

It is a moot point as to whether the two diagnoses are assessments or evaluations but assessment is probably the preferable word, since we said at the beginning of the chapter that the term would be reserved for things to do with individuals. Irrespective of this, the distinction between norm-referenced and criterion-referenced holds.

**Norm-referenced assessment**

In a norm-referenced assessment a test is used to find the performance of an individual on a defined set of tasks in relation to the performance of others who complete the same test.

> *For example, a test in fitting and machining that consisted of twenty short answer questions which are known to vary in difficulty from easy to hard, when used to compare the performances of students in a class, would be a norm-referenced test.*

**Criterion-referenced assessment**

In a criterion-referenced assessment a test is used to find the performance of an individual in relation to some specified standard.

> *For example, a test of spelling for which the teacher announced that she required 9 correct answers out of 10 in order to 'pass' would be a criterion-referenced assessment. (The standard being 9 out of 10).*

**Using norm-referenced and criterion-referenced assessment**

Norm-referenced assessment provides a comparison with other individuals, criterion-referenced assessment provides a comparison between the individual and some previously defined objective or set of objectives.

Tests used in norm-referenced assessment usually produce a range of scores, that is, they tend to spread the students out. On the other hand, criterion-referenced test scores usually clump students together, the point being that spreading students over a range is not essential to a criterion-referenced test.

A criterion-referenced test is analogous to a high jump in which the high jump bar is fixed at 1 metre. Everybody taking the test is either a successful jumper or an unsuccessful jumper.

Note: it does not matter by how much a test taker clears that bar—the task is simply to get over (i.e. to pass the test).

## Comparisons between norm-referenced and criterion-referenced assessments

The following table contrasts some of the more important aspects of the two types of assessment.

| | NORM-REFERENCED | CRITERION-REFERENCED |
|---|---|---|
| USES | To provide a means of comparing the performance of individuals on a set of defined tasks.<br><br>As a selection device to select between individuals competing for scarce resources (e.g. places in an engineering school). | To determine the peformance of individuals in relation to specified criteria (objectives).<br><br>As a 'hurdle' to decide whether an individual has gained enough knowledge and/or skills (i.e. achieved sufficient mastery) to proceed to the next level of instruction. |
| CONTENT | Content can be defined by general objectives.<br><br>Test content does not necessarily match particular classroom objectives closely.<br><br>Relationship between test items and classroom goals may be indirect.<br><br>Test items are normally a representative sample from a larger pool, the assumption being that performance on the sample is an accurate indication of performance on the whole pool. | Content must be defined by specific (behavioural) objectives. Furthermore, in many TAFE subjects, performance criteria and conditions under which performance must be accomplished also need to be specified.<br><br>Test content must match classroom objectives closely.<br><br>Relationship between test items and classroom goals must be direct.<br><br>All stated objectives need to be tested by items; sampling of objectives is not appropriate. |
| QUESTION TYPE | Items which discriminate between individuals are desirable as there is usually a requirement to compare the performance of different students. | Items must match the objectives; items which discriminate between more able and less able students are not sought. Items are required to provide information about what a student can or cannot do. |

| | NORM-REFERENCED | CRITERION-REFERENCED |
|---|---|---|
| SCORES | Variability in scores obtained by the group of students taking test is desirable.<br><br>Scores provide information on relative standing of student in group and give indication of what student can do. High scores can reflect excellence of performance. | Variability in scores is irrelevant.<br><br>Scores reflect accurately what students can or cannot do. Identifying excellent performance is often not appropriate. However, as with norm-referenced tests, it is usually possible to identify performance well above the specified standard or criteria. |
| GRADES AND CRITERIA FOR SUCCESS | Variability of scores enables allocation of grades.<br><br>Scores (in some tests) can be related to performance of larger group of students giving normative data, as in a standardised test.<br>*e.g. 'He came top of the class'*<br>*'His score puts him in the 90th percentile of the population'*<br><br>Different criteria for success are possible.<br>*e.g. 50% may be classified as a pass but students may be required to score 75% if they are to proceed to next level of study.*<br>(N.B. such criteria are arbitrary) | Grades, except pass/fail, may not be appropriate.<br><br>Scores usually expressed in such a way as to reflect performance against criterion<br>*e.g. 'She can type 40 words per minute without making more than 2 errors per page'.*<br><br>Criteria for success are often matters for judgment. In many TAFE subjects such judgments are usually related to occupational performance data. In the above example more than 2 errors per page might be defined as failure to achieve the criteria. |
| RELIABILITY AND VALIDITY | A well-developed set of psychometric techniques is available for deciding on the reliability and validity of the norm-referenced test. | Test theorists are still in the process of developing generally accepted statistical procedures for determining the reliability and validity of the criterion-referenced tests. Some of the procedures currently in use involve complex mathematical formulae. |

[After Boehm (1973)]

## The future of norm-referenced tests and criterion-referenced tests

Both norm-referenced and criterion-referenced tests have their problems, but it seems unlikely that one or the other will disappear from the education scene in the future.

Advocates of the two test forms often fail to acknowledge the imperfections of their tests. This has sometimes made it difficult to decide which type of test is the better or more appropriate. For instance, although few would deny the value of specifying the objectives of teaching, arguments do start when we try to decide how specific the objectives should be.

The 'technical education' sector of TAFE has long been associated with very specific objectives and criterion-referenced testing. Indeed, this association pre-dates the advent of the term 'criterion-referenced' into the literature on testing.

Whenever the purpose of assessment is to measure the performance of an individiual on a job-related task, a criterion-referenced test is called for. Once the performance or task to be accomplished is identified, the criterion-referenced test developers construct their test items using a knowledge of the standard of performance required and the conditions under which that performance must be accomplished. These points are taken up in more detail in the next section.

Norm-referenced testing has more relevance when issues of selection are involved. If the task is to select those individuals in a group who are most likely to succeed in a course, and there are few places available, then criterion referenced testing is of limited value.

> *For example, if 20 students have met the basic criteria (i.e. passed the criterion-referenced test) but there are only 5 places available for the 20, selection is a problem. In such cases a norm-referenced test which spreads the students out over a range of scores can make selection easier—places are simply offered to the individuals with the five highest scores.*

Sometimes it is possible to modify criterion-referenced testing procedures to have a base criterion (a pass) and progressively graded performances above the basic criterion.

> *An example could be machining a rod to a particular set of tolerances and having three progressively more demanding sets of tolerances for which the student could be rewarded. This would be an acceptable way of distinguishing between our 20 students competing for the 5 places; however it is important to realise that we are no longer dealing with a pure criterion-referenced test situation. The act of spreading the students over a range has produced a hybrid criterion-referenced and norm-referenced test.*

**OBJECTIVES**

Objectives are the subject of the next chapter and will not be dealt with in great detail at this stage. However, objectives need to be mentioned in an introductory chapter since they are the starting point of all tests.

Criterion-referenced testing requires that the objectives of instruction be clearly defined in observable behavioural terms.

A general objective like 'understands the meaning of the term scientific hypothesis' is an unsuitable basis for a criterion-referenced test although a norm-referenced test could use such an objective.

Specific (behavioural) objectives define the content of the criterion-referenced test but there is still a further requirement and that is a definition of minimal acceptable performance.

## Minimal acceptable performance and level of specificity in criterion-referenced tests

For most purposes it is enough to specify how well the learner must perform for the behaviour to be acceptable.

This normally involves

- identifying the outcome required (end behaviour) and
- indicating how well the learner must perform for the behaviour to be acceptable (criterion).

  *For example, 'can type 40 words per minute without making more than 2 errors per page'.*

However, some people insist on a much greater degree of precision and require the specifications to also:

- state the conditions under which behaviour is to occur by:
  - indicating what materials, if any, will be provided to the learner
  - stipulating the time limits within which the learner must complete the task.

  *Consider an example of such a precise specification.*

  *Students will be able to construct and measure to an accuracy of 1 millimetre a perpendicular line from the apex to the base of a triangle in at least four out of five examples presented within a 30 minute period, using only a ruler, compass and pencil.*

Tightly prescriptive objectives like the above have drawn their share of criticism mainly because of the constraints they place on teachers.

Critics point out that levels of performance and time limitations are sometimes arbitrary and claim that teaching based on such a detailed set of requirements is over-formal, rigid and dull. However such criticisms are usually 'course specific'. In some TAFE courses, meeting these specific conditions is the challenge of the learning process. A TAFE student's success in the world of work depends often upon how accurately and how efficiently he or she performs certain tasks. These points are taken up again in the chapter on Objectives.

## REFERENCES

Boehm, A. E. Criterion-referenced assessment for the teacher. Teachers College Record, 1973, 75(1), 117-126.

Mathematics item bank, ACER, 1978.

## EXERCISE I (1)

## DIAGNOSTIC AND ACHIEVEMENT TESTS

Complete the following table by writing in the empty boxes the appropriate comparison between a diagnostic and a norm-referenced achievement test.

You can check your answers by turning to page 20.

| CHARACTERISTIC OF TEST | DIAGNOSTIC | NORM-REFERENCED ACHIEVEMENT |
|---|---|---|
| Purpose | To identify a student's specific learning difficulties. | |
| Ideal spread of marks | | Spread across a range of scores, from high to low achievement. |
| Importance of time limits | | Depends on test specifications but it is usual for the limit to be stipulated. |
| Hazards | Inappropriate use of a student's total mark. | |
| Uses | | Selection and prediction. |

## EXERCISE I (2)

## NORM-AND CRITERION-REFERENCED TESTS

Below are a number of points about norm-referenced and criterion-referenced tests.

Complete the table by writing into the empty boxes the contrasting points.

You can check your answers by turning to page 21.

| CHARACTERISTICS | NORM-REFERENCED TEST | CRITERION-REFERENCED TEST |
|---|---|---|
| USE | | As a 'hurdle' to decide whether an individual has gained enough knowledge and/or skills (i.e. achieved sufficient mastery) to proceed to the next level of instruction. |
| CONTENT | Content can be defined by general objectives. | |
| SCORES | Scores provide information on relative standing of student in group and give indication of what student can do. High scores can reflect excellence of performance. | |
| GRADES | Variability of scores enables allocation of grades. | |
| QUESTION TYPE | | Items must match the objectives, discriminating items are not sought. Items are required to provide information about what a student can or cannot do. |

## SUGGESTED ANSWERS

### EXERCISE I (1)

| CHARACTERISTIC OF TEST | DIAGNOSTIC | NORM-REFERENCED ACHIEVEMENT |
|---|---|---|
| Purpose | To identify a student's specific learning difficulties. | To measure the degree to which a particular set of abilities is present in an individual. Measured against other individuals). |
| Ideal spread of marks | Not relevant. | Spread across a range of scores, from high to low achievement. |
| Importance of time limits | Must be generous but slowness itself can be a specific difficulty. | Depends on test specifications but it is usual for limit to be stipulated. |
| Hazards | Inappropriate use of a student's total mark. | Decisions regarding achievement dependent upon comprehensiveness and quality of test items.<br>Since comparisons are being made between individuals there can be no absolute standard. |
| Uses | Identification of areas in need of remediation.<br>Making decisions about individual instruction. | Selection and prediction. |

## SUGGESTED ANSWERS

## EXERCISE I (2)

| CHARACTERISTICS | NORM-REFERENCED TEST | CRITERION-REFERENCED TEST |
|---|---|---|
| USES | As a selection device to select among individuals competing for scarce resources (e.g. places in a medical school). As a means of comparing an individual against the performance of a population (e.g. I.Q. test). | As a 'hurdle' to decide whether an individual has gained enough knowledge and/or skills (i.e. achieved sufficient mastery to proceed to the next level of instruction). |
| CONTENT | Content can be defined by general objectives. | Content must be defined by specific (behavioural) objectives. |
| SCORES | Scores provide information on relative standing of student in group and give indication of what student can do. High scores can reflect excellence of performance. | Scores reflect accurately what students can or cannot do. Identifying excellent performance is often not appropriate. |
| GRADES | Variability of scores leads to allocation of grades. | Grades, except pass/fail, may not be appropriate. |
| QUESTION TYPE | Items which discriminate between individuals are desirable as there is usually a requirement to compare the performance of different students. | Items must match the objectives, discriminating items are not sought. Items are required to provide information about what a student can or cannot do. |

## CHAPTER II

## OBJECTIVES

### WHAT THIS CHAPTER IS ABOUT

This chapter explains the importance of objectives in student assessment.

An introduction is given to various types of educational objectives.

Examples are given of the relationship between the educational objective and the test item which is written to assess that objective.

*OBJECTIVES*

*After completing this chapter, you should be able to:*

- *explain the meaning of the term 'educational objective'*
- *know that there are three types of objectives- cognitive (thinking), affective (feeling), and psychomotor (doing); and recognise each type*
- *classify particular test items in terms of the educational objectives they are measuring*
- *understand that objectives have limitations in some areas of education and explain why this is so.*

### INTRODUCTION

An **educational objective** is a planned outcome which results from some process of learning. A 'planned outcome' could be the ability to measure the diameter of a piece of wire using a micrometer screw gauge. The 'process of learning' which makes this possible could involve the student receiving instruction from a teacher and then practising the skill with various pieces of wire.

However, some educators would see a problem with the above example in that its level of specificity (detail) is in doubt. For instance, we do not know how accurate the student must be in order to be judged as having mastered the objective.

Educational objectives can, in fact, range from broad statements (called goals or aims) to very precise statements, called behavioural or specific objectives. Behavioural objectives are so called because they specify the performance (behaviour) the student needs to display to satisfy the objectives.

The levels at which objectives exist can be seen from the following example.

| | |
|---|---|
| TO APPRECIATE THE IMPORTANCE OF SCIENCE AND TECHNOLOGY TO SOCIETY | is a very broad objective of education, which is better called an AIM or GOAL |
| **whereas** | |
| TO UNDERSTAND THE MEANING AND USE OF SCIENTIFIC HYPOTHESES | while still a broad statement, is in a subset of the above and can be called a GENERAL OBJECTIVE |
| **but** | |
| TO DEFINE THE MEANING OF THE TERM 'SCIENTIFIC HYPOTHESIS' | are two examples of BEHAVIOURAL OBJECTIVES arising from the general objective above. (Note that they specify what the student must be able to do). |
| **and** | |
| TO DISTINGUISH BETWEEN CORRECT AND INCORRECT SCIENTIFIC HYPOTHESES | |

Some educators increase the level of specificity even more by adding the **conditions** under which the behaviour is to occur, and the minimally acceptable **performance levels** as well.

> *For example, in Chapter I a highly specific objective was quoted as:*
>
> > *Students will be able to construct and measure to an accuracy of 1 millimetre, a perpendicular line from the apex to the base of a triangle, in at least four out of five examples presented within a 30 minute period, using only a ruler, compass and pencil.*

As objectives become more and more specific they begin to read like actual tests. Indeed, some authors refer to 'performance objectives' when describing any objective which specifies:

- a **skill to be performed**
- the **conditions** under which it must be performed
- the **acceptable standard** of performance to be achieved.

Such an approach is relevant to many areas of TAFE, especially when the aim is to relate a performance test to the real world.

The example of measuring the diameter of wire can be turned into a performance objective by prescribing the acceptable tolerance limits within which the measurement must be made.

*PERFORMANCE OBJECTIVE*

| | |
|---|---|
| *Skill* | *measure the diameter of a copper wire* |
| *Conditions* | *given a length of wire and micrometer screw gauge* |
| *Standard* | *to an accuracy of + or - 0.005 mm.* |

## CLASSIFICATION OF OBJECTIVES

There are a number of systems for classifying educational objectives, however, this Chapter will consider only the scheme devised by Bloom and his co-workers (1956).

> *This system has its critics and is not being presented here as an exemplary method of classification. It is, however, the best known and most researched of all the methods and a major advantage is that over the years it has been used to classify many thousands of test items.*
>
> *TAFE teachers should see it as a useful method of classification but should feel free to modify it as necessary to meet their needs.*

The system developed by Bloom and his co-workers is called a taxonomy because it is a system of classification. The taxonomy classifies within three basic areas or domains, namely:

- cognitive
- affective
- psychomotor.

The **cognitive domain** includes objectives which deal with the recall or recognition of knowledge and the development of intellectual abilities and skills.

The **affective domain** objectives describe changes in interests, attitudes and values, and the development of appreciation.

The **psychomotor domain** deals with manipulative (motor) skills.

These three domains are sometimes referred to as the 3H's; Head (cognitive), Heart (affective) and Hands (psychomotor) to help emphasise the distinctions.

*COGNITIVE* = *THE HEAD* = *THINKING*

*AFFECTIVE* = *THE HEART* = *FEELING*

*PSYCHOMOTOR* = *THE HANDS* = *DOING.*

Each domain is broken down into a number of levels as follows:

Cognitive Domain

- Knowledge
- Comprehension
- Application
- Analysis
- Synthesis
- Evaluation

} levels of **cognitive** learning

Affective Domain

- Receiving
- Responding
- Valuing
- Organisation
- Characterisation

} levels of **affective** learning

Psychomotor Domain

- Reflex movements
- Basic-fundamental movements
- Perceptual abilities
- Physical abilities
- Skilled movements
- Nondiscursive communication

} levels of **psychomotor** learning

A brief outline of the COGNITIVE DOMAIN levels:

i. **Knowledge**—the recall of specific or general information, patterns, structures, or settings.

*The ability to recall the names of different tools involves 'knowledge'.*

ii. **Comprehension**—the level of understanding at which an individual knows what is being communicated and can use this knowledge in some way.

There are actually three types of comprehension.

The first involves **translation**, that is, putting a communication into another language.

*Reading a graph involves translation.*

The second type is **interpretation.**

*Reading a paragraph and being able to identify and comprehend the major ideas involves interpretation.*

**Extrapolation** is the third type.

*Predicting a trend based on a comprehension of the information provided involves extrapolation.*

iii. **Application**—the use of abstractions in particular and concrete instances.

*The ability to predict the probable result of introducing a change such as substituting one metal for another in the manufacture of batteries involves 'application'.*

iv. **Analysis**—the breaking down of a communication into its component parts and understanding of the relationship between these parts.

*The ability to read a report and distinguish fact from opinion requires 'analysis'.*

v. **Synthesis**—the arrangement and combination of pieces, parts, or elements into a coherent pattern or structure.

*The skill in writing a report that requires the organisation and development of ideas involves 'synthesis'.*

vi. **Evaluation**—the quantitative and qualitative value judgments on factors such as materials, methods, the use of a standard of appraisal, and the utilisation of criteria.

*The ability to compare two conflicting ways of carrying out a project and judge the advantages and disadvantages of each requires 'evaluation'.*

Bloom et al (1956) can be consulted for a more detailed breakdown of the levels i to vi.

## Cognitive domain examples

The examples given below are meant to help the reader understand the Bloom taxonomy. It should be noted that classifying items, other than knowledge items, in this way assumes the particular answer is not arrived at by recalling something already taught. What can involve 'comprehension' or 'application' for some students may be 'knowledge' for others.

### Knowledge

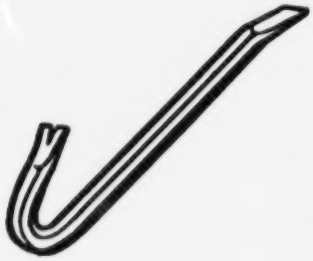
*K1 The tool shown in the diagram is*

*A a stake*

*B a bossing mallet*

*C a pinch bar*

*D an anvil*

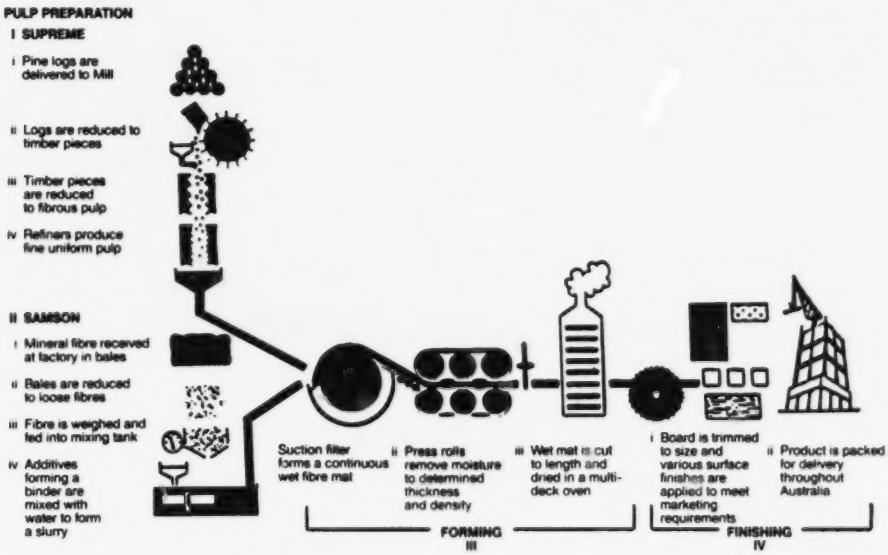


(Answer C)

This item tests knowledge of a specific piece of information, that is, knowing the name of a particular tool.

### Comprehension

For questions C1 and C2 consider the diagram which describes the process of fibre board manufacturing.


PULP PREPARATION
I SUPREME
i Pine logs are delivered to Mill
ii Logs are reduced to timber pieces
iii Timber pieces are reduced to fibrous pulp
iv Refiners produce fine uniform pulp
II SAMSON
i Mineral fibre received at factory in bales
ii Bales are reduced to loose fibres
iii Fibre is weighed and fed into mixing tank
iv Additives forming a binder are mixed with water to form a slurry
Suction filter forms a continuous wet fibre mat
ii Press rolls remove moisture to determined thickness and density
iii Wet mat is cut to length and dried in a multi-deck oven
i Board is trimmed to size and various surface finishes are applied to meet marketing requirements
ii Product is packed for delivery throughout Australia
FORMING III
FINISHING IV



BEST COPY AVAILABLE

C1 *Each stage in the process of manufacturing Supreme and Samson boards has been identified by a number. For example, I ii is the stage at which the 'Logs are reduced to timber pieces'.*

*Which of the following most accurately identifies the stages at which water is removed from the material being processed?*

A *I ii, II iv, III ii, III iii, IV i*

B *II iv, III ii, III iii, IV i*

C *II iv, III ii, III iii*

D *III ii, III iii.*

(Answer D)

This item requires the identification and translation of information presented in diagrammatic form. It is an example of the Translation level of a Comprehension question.

C2 *Which of the following best describes the processes of manufacturing Supreme and Samson board?*

*The processes for manufacturing Supreme and Samson*

A *are similar in most respects*

B *are similar but differ mainly in the pulp preparation stage*

C *are similar but differ mainly in the finishing stage*

D *are different in most respects.*

(Answer B)

This item requires the interpretation of a communication given in the form of a diagram. It is an example of the Interpretation level of a Comprehension question.

**Application**

Teachers and curriculum designers have long recognised that students do not really 'understand' an idea or principle unless they can **apply** it in a new problem situation. Success with 'application' questions, therefore, implies the student has understood the topic concerned.

Consider the following item:

*APP1 A straight stick half in and half out of a bowl of water appears bent because*

*A light rays do not travel in a straight line in water*

*B water droplets 'bend' the light rays*

*C light rays change direction slightly when they go from water to air*

*D light rays scatter when they enter water.*

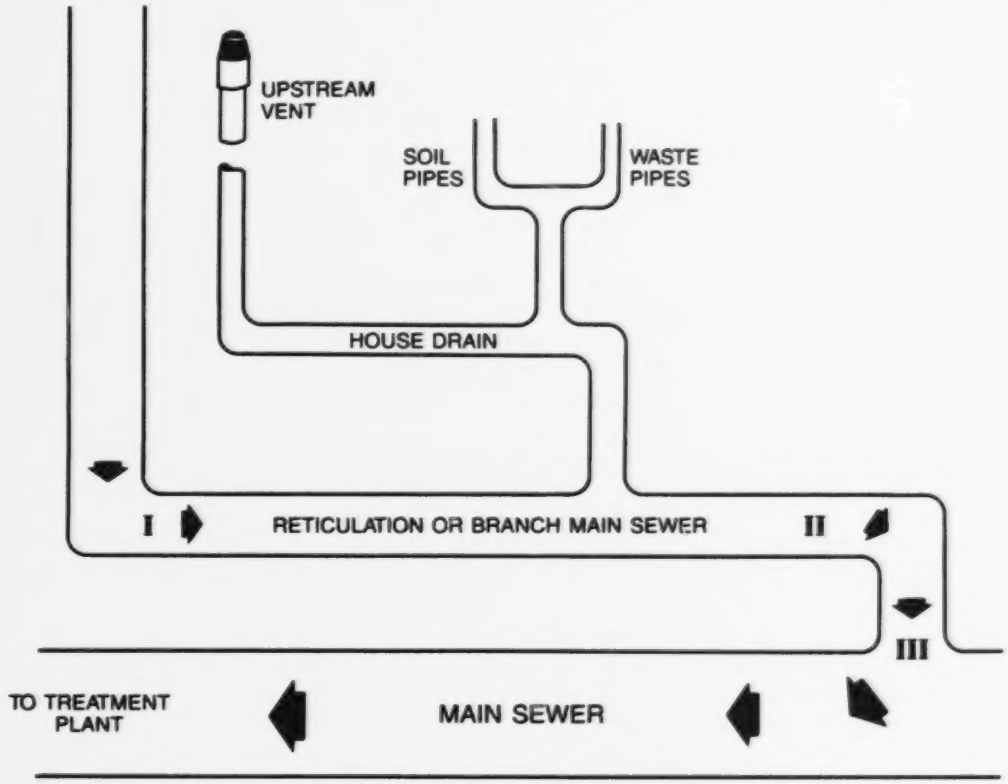
(Answer C)

This item requires the student to explain a new phenomenon—the apparent bending of the stick—in terms of a known principle, that is, the diffraction of light.

Note:

In order to classify APP1 as an 'application' question it is necessary to assume the student has not been taught the answer. If the reason for the stick appearing bent has been taught then this becomes a 'knowledge' question.

For APP2 consider the following diagram of a sewerage system.




**SEWERAGE SYSTEM**

*APP2 If faults or blockages in the system can occur at points I, II and III, which of these is the responsibility of a plumber employed by a householder*

*A all of I, II and III*

*B I and II only*

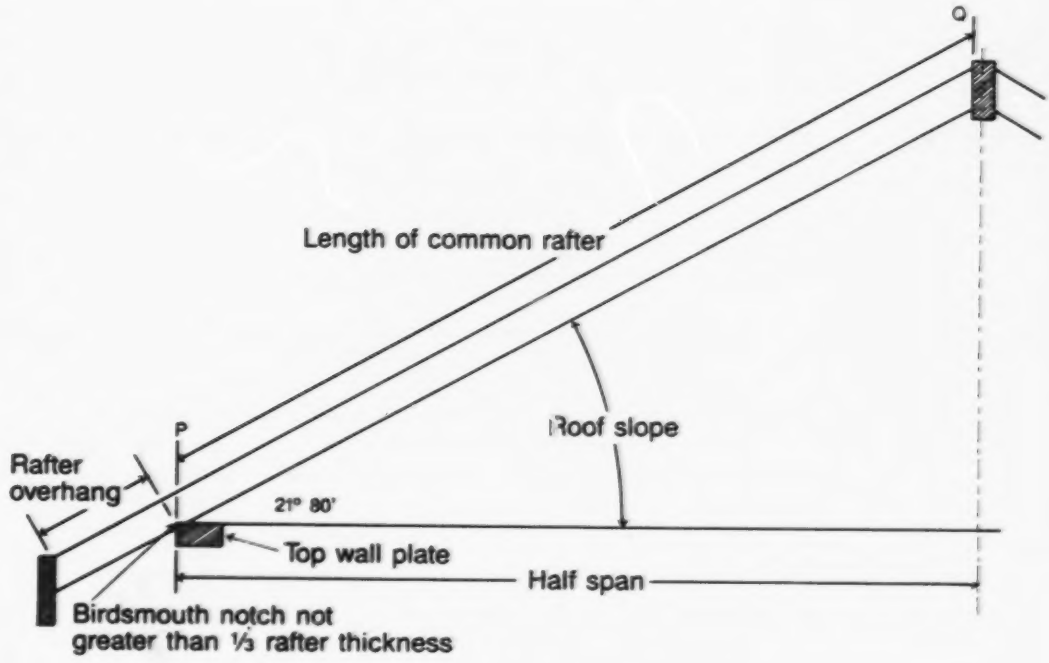
*C I only*

*D none of I, II or III.*

(Answer D)

This item requires the application of a general rule or procedure; it is also arguably an 'interpretation' question, however, the rule of thumb for classification is to accept the higher level.

*APP3 In the following diagram, if the length of PQ is 3 metres and the roof slope is 21° 80', what is the length of the half span?*




**COMMON RAFTER**

This item tests the ability to apply the laws of trigonometry to a practical situation.

APP4 *Enter the following transactions into the journal and ledgers (general and subsidiary) of J. Chaston. Balance the accounts, prepare a trial balance and debtors' and creditors' schedules at 31 May 1984.*

*J. CHASTON*

*1984*
*May*

*1 Commenced business with the following assets and liabilities:*
*Debtors $700; Buildings $4,500; Stock $540; Bills receivable $400; Office furniture $370; Creditors $580; Bank overdraft $800; Bills payable $100*
*Debtors - W. Hall $250, W. Lofts $175, J. Gill $275*
*Creditors - G. Collis $320, C. Stewart $260*

*3 Cash sales $140 (CRS)*
*Credit purchase from B. Gill $300 (Inv.91)*

*5 Gill charged $10 cartage on purchase on Invoice 91*
*Paid advertising $60 (Ch. 31)*

*7 Cash sales $640 (CRS)*
*Loan from R. Crowe $1,000 - Received*

*9 Credit sales to J. Miller $250 (Inv. 21) and B. Cox $145 (Inv.22)*
*Proprietor drew cash for own use $50 (Ch. 32)*

*10 Goods returned to us by Miller $25 (Correction note 39)*
*Cash sale to G. Reveruzzi $440*
*Paid insurance $75 (Ch. 33)*

*15 Credit purchases from Collis $100 (Inv. 17) and P. Barry $85 (Inv. 31)*

*18 J. Gill paid us $143 and we allowed him $7 discount (Rec. 41)*
*Paid office expenses $55 (Ch. 34)*
*We returned goods to Collis $15 (C/N 40) and Barry $10 (C/N 71)*

*23 Bought furniture for office from J. Bennett for $90 cash (Ch. 35)*
*Credit sale to J. Gill $70 (Inv. 23) and J. Goold $95 (Inv. 24)*

*25 Allowance to Goold for overcharge $15 (C/N 40)*
*Paid interest on loan for May $5 (Ch. 38)*
*Hall declared bankrupt and can pay only 50 cents in $- balance written off (Rec. 42)*

*28 Cash purchases from J. Greenwood $80 (Ch. 37) and S. Silvan $200 (Ch. 38)*
*Cash sales $475 (CRS)*

*30 Paid Collis $275 and he allowed us $15 discount (Ch. 39)*
*Cox settled his debt in full by paying us $138 (Rec. 43)*

This question requires the application of principles of accounting to a new or unique situation.

**Analysis**

**Analysis** emphasises the breakdown of information into its constituent parts and the detection of the relationships of the parts and the way they are organised.

*AN1 Light has the characteristics of a bullet and a wave. Which one of the following is evidence of the wave characteristics of light?*

*A light can be reflected by a mirror*

*B a beam of light spreads out after passing through a small opening*

*C light causes a current to flow in a photo-electric cell*

*D light carries energy*

(Answer B)

This item requires the student to infer a relationship between a physical property—wave motion—and a condition that must result—diffraction.

Question AN2 refers to the following information

*CLASS C FIRES*

*CLASS C fires involve live electrical equipment such as light fittings, motors, generators, cables, wiring, switches, switchboards and electronic equipment.*

*Blanketing, or smothering the fire to exclude oxygen is the most effective way of extinguishing Class C fires.*

*Vapourising liquid (Bromochlorodifluoromethane, BCF), dry powder chemical and carbon dioxide ($CO_2$) extinguishers are effective in putting out Class C fires.*

> *WARNING:*
>
> *The vapourising liquid gives off toxic gases when in contact with fire. Take care with such equipment when used in confined spaces. Wear proper equipment, or ensure proper ventilation.*
>
> *The extinguishing agent used must be a non-conductor of electricity in order to avoid shock and further damage to equipment.*
>
> *Never use water, or foam extinguishers, on Class C fires.*

AN2 *Assuming you have suddenly come upon a fire which is burning strongly in a large switchboard attached to the wall in the switchboard room of a block of offices.*

*You have at your disposal only four items:*

*I* a fire blanket

*II* a BCF extinguisher

*III* a $CO_2$ extinguisher

*IV* a foam extinguisher.

*Based on the information provided, which of the following would be the most appropriate to extinguish the fire?*

*A* open doors and windows and use I

*B* close doors and windows and use II

*C* open doors and windows and use II or III

*D* close doors and windows and use IV.

(Answer C)

This item requires the student to understand and analyse the interrelationships between the points made in the information supplied.

**Synthesis**

Synthesis is the putting together of parts to make a whole.

Most essay questions fall into this category—it is rare for an objective multiple choice question to qualify for the synthesis level.

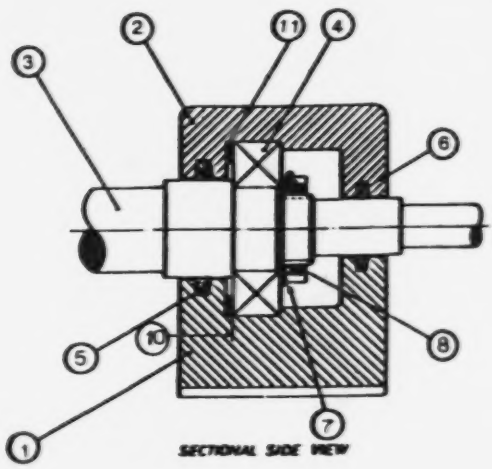
*S1 Compare and contrast the first aid procedures you would use in treating a person who had been badly burnt on the arm by hot acid and was at the same time unconscious as a result of inhaling the acid fumes.*

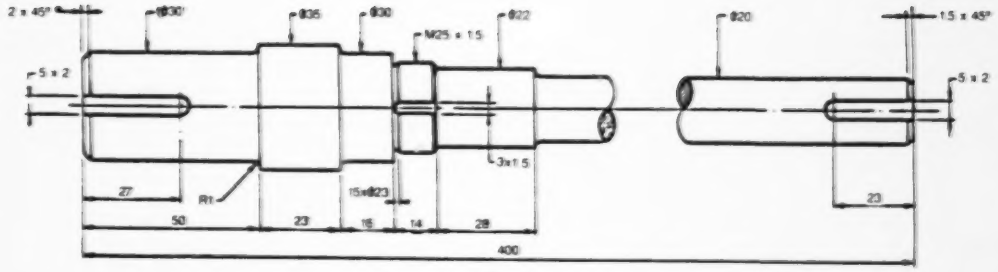
This item tests the ability of the students to develop a plan of action to satisfy two sets of conditions.

*S2 Below are partially completed drawings of a split bearing house.*

*Referring to the sectional side view and shaft detail, and to data on relevant parts, complete the drawings by adding a third view showing all internal details of the housing only.*

*Completely dimension the drawing. Label third view and show any missing details, dimensions, cutting planes and surface texture symbols where required.*






SHAFT

| | | | |
|---|---|---|---|
| 11 | SPACING WASHER | 1 | MS |
| 10 | SCREWS SOC HDCAP M6 x 1 x 25 | 4 | STD |
| 9 | DOWELS Ø 6 x 40 | 2 | STD |
| 8 | LOCKING NUT CAT. N2 | 1 | STD |
| 7 | LOCKING WASHER CAT. N2 | 1 | STD |
| 6 | SEAL CAT. N2 | 1 | STD |
| 5 | SEAL CAT. N2 | 1 | STD |
| 4 | BEARING DEEP GROOVE CAT. N2 | 1 | STD |
| 3 | SHAFT Ø 35 x 400 | 1 | MS |
| 2 | BEARING TOP | 1 | MS |
| 1 | BEARING BASE | 1 | MS |
| ITEM | DESCRIPTION | QTY | MATL |
| SPLIT BEARING ASSEMBLY | | | |
| DRN. J.L. | CKD. P.B. | APPD. R.K. | DRAWING NO. PB - 7 |

This item requires the development of a plan to satisfy a set of prerequisite conditions.

**Evaluation**

Evaluation is about making judgments. It can involve some combination of all the other behaviours of Knowledge, Comprehension, Application, Analysis and Synthesis. What is added are criteria involving values. Evaluation is a link between the cognitive and affective domains but with the emphasis on the cognitive.

E1 *It may be worthwhile for the manufacturer of business goods to use space in magazines specialising in current affairs in order to have the opportunity to reach those readers who as business people are interested in buying the manufacturer's product. Consider each of the following business goods. Tick (✔) box A or B opposite each product as follows:*

A *if such an approach would appear to provide enough worthwhile coverage to justify the expenditure;*

B *if such an approach would not be likely to reach enough good prospects to justify the expenditure.*

| | A | B |
|---|---|---|
| *Holden car* | ☐ | ☐ |
| *National rock drills (mining industry)* | ☐ | ☐ |
| *IBM typewriters* | ☐ | ☐ |
| *Buckner textile machines* | ☐ | ☐ |
| *Xerox Bond Paper.* | ☐ | ☐ |

This item involves judging an approach to advertising in terms of a particular criterion.

E2 *A home owner has a room with a stained pine floor which is in good condition. She needs to decide between two floor treatments—polishing with a floor wax or painting with a clear epoxy resin.*

*For each of the qualities listed below check (✔) which treatment is superior. If there is no difference between treatments do not check a box.*

| *Quality* | *TREATMENT* | |
|---|---|---|
| | *Polished* | *Painted* |
| *resistance to scratching* | ☐ | ☐ |
| *resistance to sunlight* | ☐ | ☐ |
| *overall appearance* | ☐ | ☐ |
| *safety when walking over surface* | ☐ | ☐ |
| *frequency of treatments over year* | ☐ | ☐ |
| *cheapness of treatment* | ☐ | ☐ |

This item involves making a number of judgments against a set of given criteria (qualities).

The two sections which follow introduce the different sub-divisions of the Affective and Psychomotor Domains. No items are provided in these sections. Chapter III, however, contains a number of items from these two domains.

**Affective Domain**

The sub-divisions of the affective domain are:

i. Receiving—the sensitivity of the learner to the existence of certain stimuli. This level of affective learning includes three categories:

   (a) awareness

   (b) willingness to receive

   (c) controlled or selected attention.

ii. Responding—the active response of the student to stimuli. This level of affective learning has three sub-categories:

   (a) acquiescence to respond

   (b) willingness to respond

   (c) satisfaction in response.

iii. Valuing—the development of an attitude toward a stimulus by the student. The process of valuing is further divided into:

   (a) acceptance of value

   (b) preference for a value

   (c) commitment to a value.

iv. Organisation—the level of affective learning at which the student begins to build a value system. This category is subdivided into:

   (a) conceptualisation of a value

   (b) organisation of a value system.

v. Characterisation of a Student's Value Complex—the organisation of a student's values into some kind of internally consistent system. This level of affective learning is divided into:

   (a) the development of a generalised set

   (b) the characterisation of a value system having as its object all that is known or knowable.

**Psychomotor domain**

The subdivisions of the psychomotor domain are:

i. Reflex movements—involuntary movements of the body which are functional at birth and develop with the maturation of the individual.

ii. Basic-fundamental movements—inherent movement patterns which form the basis for specialised complex skilled movements.

iii. Perceptual abilities—ways in which stimuli are responded to by the brain centres when the learner is making a response decision.

iv. Physical abilities—the functional characteristics used when an individual makes skilled body movements.

v. Skilled movement—the result of acquisition of a degree of efficiency when performing a complex movement task.

vi. Nondiscursive communication—communication by body movements rather than by spoken or written language.

**USEFULNESS OF BEHAVIOURAL OBJECTIVES**

Educational objectives specify the planned outcomes of some process of learning. As such they identify the purpose of instruction to both teacher and student.

The main values of clearly stated, unambiguous objectives are that they enable:

- the teacher to know what content to teach and to what level of detail
- the students to know what they are expected to be able to do following instruction
- the test developer (usually the teacher) to know what type of questions are appropriate to assess whether or not the students have mastered the objectives.

In most TAFE courses objectives can be partitioned into three components, namely:

- performances
- conditions
- standards.

These three components are variously described by other synonyms in the literature as reported by Swezey (1981) in Table 1.

**Table 1**

**SYNONYMS FOR COMPONENTS OF AN OBJECTIVE**

| Performance | Conditions | Standard |
|---|---|---|
| . Activity<br>. Action<br>. Skill<br>. Knowledge<br>. Response<br>. Task<br>. Reaction | . Test condition<br>. Required equipment<br>. Required facilities<br>. Required personnel<br>. Required materials<br>. Environment<br>. Set-up<br>. Job condition | . Criterion<br>. Cut-off<br>. Passing grade<br>. Requirement<br>. Go, no-go standards |

**LIMITATIONS OF BEHAVIOURAL OBJECTIVES**

The following quotation is taken from an article by Lacey and Lawton (1981) in 'Issues in Evaluation and Accountability'. It sums up the substantive criticisms of the objectives approach to education. Put most simply the authors are saying objectives take the art and heart out of teaching.

> The behavioural objectives approach rests on the assumption that for all teaching and learning situations it is possible and desirable to specify in advance the precise changes which will be observed (and measured) in pupil behaviour. The essence of a behavioural objective is that it must be a pre-specified change in pupil behaviour which is testable and measurable. One major objection to this is that in some kinds of learning situations it is neither possible nor desirable to predict pupil responses: in an English literature lesson, for example, the teacher must have ideas about the kind of response which he hopes for as well as those he would find disappointing, but these are not precise changes in behaviour—each individual's response to a work of art is unique. Similarly, in a history lesson, there is often no one right answer to a question—a pupil's response has to be rated according to several different criteria such as knowledge of the evidence, ability to evaluate the evidence and draw conclusions from it, ability to argue logically, etc. These may be regarded as general objectives, but they certainly do not qualify as precise behavioural objectives. Many would argue, however, that without such practices education would be greatly impoverished. (p. 230)

The reader's attention is drawn, in particular, to the relevance of the above criticisms to the problem of measuring attainment in the affective domain. Objectives related to such things as tolerance and community spirit can be likened to the history example. Reducing such things to observable behavioural changes may distort and trivialise the whole educational process.

These limitations apply in varying degrees across TAFE. They are most valid in the traditional 'arts' subjects like English and history and also in TAFE subjects which have aesthetic components such as drawing and design. However, the limitations referred to apply more to teaching than testing. There is no need for testing objectives to be constraining. The percipient English teacher can, for example, test in such a way that unique responses to essay questions can be suitably rewarded.

Clearly many TAFE subjects are, of necessity, structured around sets of performance objectives which involve varying mixes of cognitive, affective and psychomotor skills. What TAFE teachers must appreciate is that objectives need to be considered as a part of the educational process. Their usefulness is intimately bound up in such things as the subject matter and teaching approach.

## REFERENCES


Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. Taxonomy of educational objectives: handbook 1 cognitive domain. London: Longman. 1956.

Lacey, C., & Lawton, D. (Eds.). Issues in evaluation and accountability. New York: Methuen. 1981.

Swezey, R. W. Individual performance assessment: an approach to criterion-referenced test development. Reston, Virginia: Reston. 1981.

## EXERCISE II (1)

## CLASSIFICATION OF OBJECTIVES

Listed below are fifteen educational objectives.

Identify the domain to which each belongs by writing the letter C, A or P alongside (cognitive = C; affective = A; psychomotor = P).

| | OBJECTIVE | DOMAIN |
|---|---|---|
| 1. | Displays proficiency in the use of refrigeration, engineering and other hand tools | _____ |
| 2. | Keeps work place neat and tidy | _____ |
| 3. | Calculates swept volume of a car cylinder given stroke length and diameter of bore | _____ |
| 4. | Writes names of all parts of an electric motor | _____ |
| 5. | Writes comprehensive reports based on field observations | _____ |
| 6. | Completes additional non-compulsory assignments | _____ |
| 7. | Solders together two sheets of galvanised iron after being shown once | _____ |
| 8. | Estimates the costs of labour and materials involved in constructing a three bedroom house given plans and specifications | _____ |
| 9. | Completes jobs in a methodical and efficient manner | _____ |
| 10. | Judges the comparative effectiveness of alternate solutions to practical problems of land use | _____ |
| 11. | Plans menus to meet pre-determined budget requirements | _____ |
| 12. | Prepares company financial statement for inclusion in published annual reports | _____ |
| 13. | Uses the tools of the trade with perfect ease | _____ |
| 14. | Works effectively without direct supervision | _____ |
| 15. | Mixes paints to match faded colours of automotive vehicles so that they are indistinguishable to the unaided eye. | _____ |

## EXERCISE II (2)

## CLASSIFICATION OF QUESTIONS BY OBJECTIVE

Read the following questions and classify each one according to the cognitive objective you believe it is testing.

1. The most suitable method of joining aluminium sheeting and tube is by means of:

   A Fusion welding
   B Brazing
   C Argon arc welding
   D Soft soldering.

   (Answer C)

   Objective _____

2. If an electric refrigerator is operated with the door open in a perfectly insulated sealed room, what will happen to the temperature of the room?

   A It will rise slowly
   B It will remain constant
   C It will drop slowly
   D It will drop rapidly.

   (Answer A)

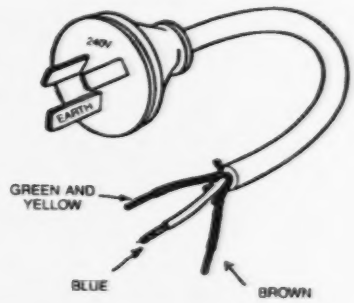
   Objective _____

3. Three wires in a lead from an electric plug are 'colour-coded' as shown in the diagram. Which wire comes from EARTH?

   A Brown
   B Blue
   C Green and Yellow
   D Either Blue or Brown but not green and yellow.

   (Answer C)

   Objective _____




INTERNATIONAL THREE CORE LEAD CODING

4. The discs in the human spine are easily damaged if they are stressed while the back is bent. The spine resists damage best when it is held straight and rigid so that it is braced by the back and stomach muscles. The most powerful muscles that can be used for lifting are in the thighs and buttocks.

Based on the information provided, which of the diagrams **best** represents an incorrect lifting position?

A I only

B II only

C III only

D I, II and III.

(Answer A)

Objective _____

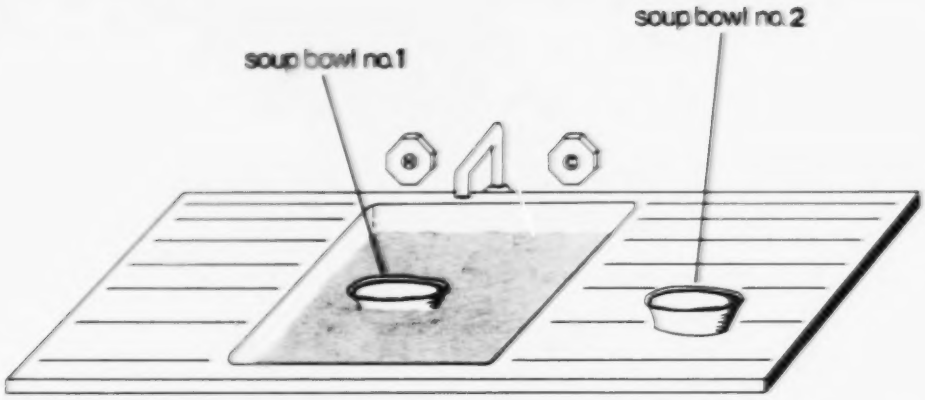
I

II

III

5. Explain in note form the sequence of steps you would carry out to revive an injured person who has a heartbeat but is not breathing.

Objective _____




6. A soup bowl is floating in a kitchen sink as shown in the diagram.

If a second bowl is placed inside the first, what happens to the level of water in the sink?

A It drops slightly

B It rises slightly

C It remains the same

D A, B or C can be correct as it depends upon the size of the second bowl.

(Answer B)

Objective _____

## SUGGESTED ANSWERS—EXERCISE II (1)

| | OBJECTIVE | DOMAIN |
|---|---|---|
| 1. | Displays proficiency in the use of refrigeration, engineering and other hand tools | P |
| 2. | Keeps work place neat and tidy | A |
| 3. | Calculates swept volume of a car cylinder given stroke length and diameter of bore | C |
| 4. | Writes names of all parts of an electric motor | C |
| 5. | Writes comprehensive reports based on field observations | C |
| 6. | Completes additional non-compulsory assignments | A |
| 7. | Solders together two sheets of galvanised iron after being shown once | P |
| 8. | Estimates the costs of labour and materials involved in constructing a three bedroom house given plans and specifications | C |
| 9. | Completes jobs in a methodical and efficient manner | A |
| 10. | Judges the comparative effectiveness of alternate solutions to practical problems of land use | C |
| 11. | Plans menus to meet pre-determined budget requirements | C |
| 12. | Prepares company financial statement for inclusion in published annual reports | C |
| 13. | Uses the tools of the trade with perfect ease | P |
| 14. | Works effectively without direct supervision | A |
| 15. | Mixes paints to match faded colours of automotive vehicles so that they are indistinguishable to the unaided eye. | C |

[Adapted from New South Wales Department of TAFE. Developing a TAFE syllabus using aims and objectives. Sydney: undated (pp. 97-98).]

## SUGGESTED ANSWERS—EXERCISE II (2)

| QUESTION | COGNITIVE OBJECTIVE |
|---|---|
| 1 | KNOWLEDGE (of methodology) |
| 2 | COMPREHENSION (interpretation) |
| 3 | KNOWLEDGE (of conventions) |
| 4 | COMPREHENSION (translation) |
| 5 | SYNTHESIS |
| 6 | APPLICATION. |

# CHAPTER III

# TYPES OF QUESTIONS

## WHAT THIS CHAPTER IS ABOUT

This chapter has two purposes—to describe briefly some of the types of questions commonly used in classroom tests and to make you aware of the particular features of each type. Each question type is presented in a way that enables you to know under what circumstances it should be used and to know what factors to consider when the question is being prepared.

*OBJECTIVES*

*After completing this chapter you should be able to:*

- *describe the types of questions most commonly used in assessing student learning*
- *predict which type of question is most appropriate for assessing a particular learning outcome*
- *know the guidelines for preparing the different types of questions.*

## TESTS, QUESTIONS AND ITEMS

Tests are usually made up of a number of individual questions or items which must be answered in a specified time. In this chapter the terms **test question** and **test item** will be used to mean the same thing, although the term **item** has become the more commonly accepted.

## DIFFERENCES IN OBJECTIVITY OF TESTS

The more 'objective' the test question, the less judgment that needs to be exercised in deciding if the answer to that question is correct.

A tick (✓) in the box [T], rather than box [F], to indicate the answer to a TRUE/FALSE question is TRUE provides a situation which requires very little judgment on the part of the teacher. Tests made up of TRUE/FALSE items are therefore highly objective.

The difference in 'objectivity' of different test types can be seen in Figure 1.

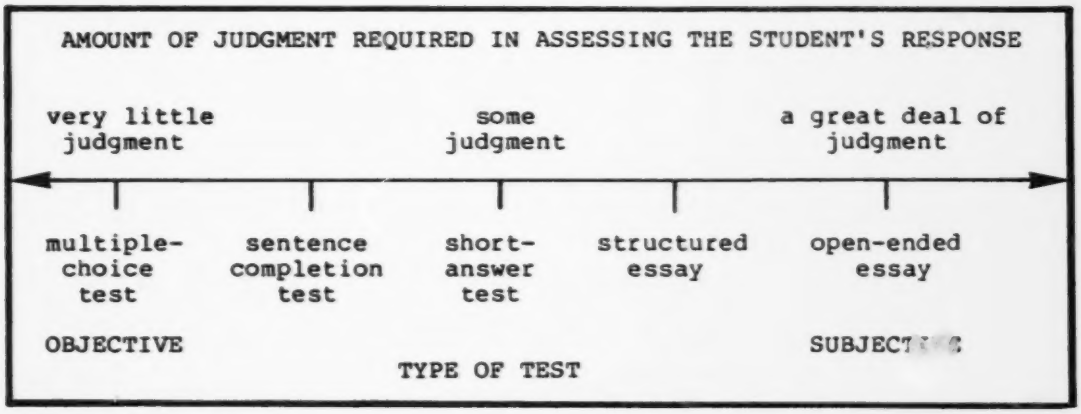



**Figure 1**

The diagram makes one point particularly clear. Scoring each type of test requires different amounts of judgment by the marker.

It is important to appreciate that, when an overall view of testing is taken, no one test type has any special advantage over another. As will be shown later, deciding what is best depends on the circumstances.

The time required to prepare and mark a particular test is often an important consideration in what types of items are used.

## MAKING GOOD USE OF AVAILABLE TIME AND RESOURCES

Too often teachers do not use the most appropriate test items because either they do not have the necessary skills to produce good items, or they worry about the time it will take to prepare a good test. However, teachers usually do have enough time to make assessments, their problem is more one of knowing how to use the time to the best advantage.

> *Consider this example. A teacher wishes to check that his or her students are able to recall a series of facts, such as dates, names, formulae, etc. There is little doubt that the most efficient way of making this check is to use a set of objective tests items (e.g. true/false or multiple-choice items). However, the writing of such items takes time and many teachers find this daunting, so they tend to use an essay or similar test.*

*The essay test, apart from taking a long time to mark, also introduces the additional problem of the teacher having to interpret exactly what the students meant by their answers. Interestingly, the additional marking time usually approximates to the time it would have taken to prepare an objective test.*

In fact when the time taken to construct a test is considered the situation depicted in Figure 1 is reversed to that shown in Figure 2.

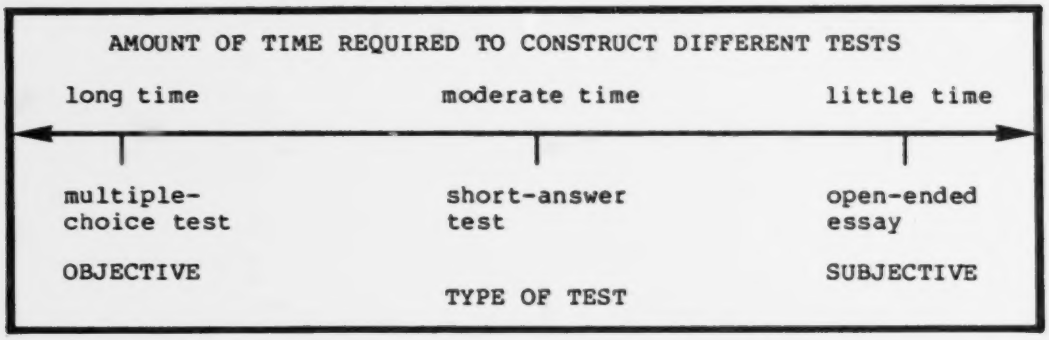



**Figure 2.**

And what is probably more important, the objective test can be reused with another class at another time. When this is done the time saved by the teacher can be considerable.

A lack of teacher skills in item writing can be partly reduced if collections of good test items are available. These collections, called **item banks**, can be used to select suitable items.

## Item banks

An item bank is a collection of questions held in a system of storage (bank) from which the items can be readily retrieved. Each item is classified in such a way that it is possible, after specifying the type of item needed, to withdraw a suitable one from the bank.

In its simplest form the bank can be a box of cards with individual items written on individual cards and suitably classified. More elaborate banks can involve the use of a computer which not only stores items but has the added advantage of maintaining item statistics. That is, the computer can print out information on individual questions, such as what percentage of students get a question correct each time it was used, and how difficult it was in comparison with other questions used in the same test.

**TYPES OF TEST QUESTIONS**

In practice, test questions do not always fall into easily identifiable categories. However, to make the following presentation clearer, the various types have been grouped together as set out below.

- **Practical Assessments**—in which the student performs some practical task based on workshop or laboratory practice.
- **Project or Field Work Assignments**—in which the student completes some long term activity resulting in a report, model, large scale construction, etc.
- **Written Assessments**—in which the student produces a written response (varying from writing an essay to marking a box) to a question based on some aspect of learning. In the jargon of the testing industry these are often called 'paper and pencil tests'.
- **Oral and Listening Assessments**—in which the students' speaking and listening skills, as they relate to an area of study, are being assessed.
- **Special Purpose Assessments**—in which novel or unusual test forms are used.

**Practical assessment**

Practical assessments involve the student performing some manipulative (hands-on) task in a college workshop or laboratory situation. Practical tests are usually done under supervision.

*Examples of practical tests*

- *Given three unknown solids A, B and C, perform experiments to determine the solubility of each in water and in alcohol*
- *Make a loaf of bread from the ingredients you have been given*
- *Use the tools and materials provided to build a wooden file box.*

Although practical tests are traditionally used to assess manual (psychomotor) skills, such as athletic skills or typing skills, their great advantage is that they can also be used to measure abilities from the cognitive and affective areas. A properly designed practical test can assess in all three domains. A full explanation of the difference between the cognitive, affective and psychomotor domains is provided in Chapter II. Put simply:

- the **cognitive** domain deals with **knowledge and intellectual abilities** (i.e. thinking)
- the **affective** domain deals with **attitudes and values** (i.e. feeling)
- the **psychomotor** domain deals with **manual** skills (i.e. doing).

Because practical assessments can be used to assess in all three domains they have much to recommend them. There are, however, particular features which require careful consideration:

**Mastery over time**

When students are required to perform experiments, observations, etc. in fixed time periods (i.e. under test conditions), this is usually an artificial situation. A single experimental test, a demonstration, etc. is not typical of real life. A cook is not judged on a single dish!

With most skills, the teaching process requires that students make several practice (training) attempts before they master a skill. Some students are slower learners than others and therefore if students are tested after, say, four practices, some will show mastery of the skill while others may still not have mastered that skill. However these 'non-masters' may eventually perform better on the skill than those students who mastered it more quickly.

Teachers should allow an adequate number of practice attempts before they finally test to determine if a student has mastered a particular skill.

Assessments can also be made on practice attempts but these should only be used for informing the teacher and the student on progress in mastering the skill.

**Objectives and record-keeping**

Practical assessment is best done on repeated occasions over a period of time and requires

- a clear statement of the skills being tested.
- maintenance of careful records showing the performance of individual students on the skills.

**Quality of workmanship and use of standards**

i Specification of quality

Where a practical exercise involves the production of a precisely defined result it is possible to establish acceptable margins of error.

*Example:* *Assessing the production of a chemical from a known starting amount.*

*After such an exercise has been completed, the yield of pure chemical is measured and marks allotted according to some predetermined rule such as:*

| *Yield of Pure Product % of possible yield* | *Number of Marks* |
|---|---|
| *98-100* | *5* |
| *90-97* | *4* |
| *80-89* | *3* |
| *70-79* | *2* |
| *60-69* | *1* |
| *below 60* | *0* |

In a similar way, a micrometer can be used to check tolerances achieved in an exercise in metal cutting and marks awarded depending on the amount of deviation from the 'ideal'.

ii Comparison with samples

Where a practical exercise involves a product that cannot be measured with such things as a micrometer or a beam balance, it may be possible for a panel of experts to agree upon a graded set of examples to guide the assessor.

**Use of checklists**

There is a need to have checklists to assess some types of practical work, and to use these periodically. These checklists should consider such things as:

i approach to the task
- ability to devise a solution, plan of the work
- initiative
- ability to give and receive instructions

ii handling equipment
- thoroughness
- accuracy
- cleanliness, tidiness

iii recording observations
- legibility
- checking
- correcting

iv translation of data
- accuracy of calculations
- preparation of report
- graphing.

62

Note that these points involve thinking (intellectual) skills and attitudes as well as manual skills. Each can be assessed using a simple five-point scale of 5 to 1 or A to E.

*For example, in assessing the way a student handles equipment when making bread, or preparing a meal the two extremes would be:*

*A checks that equipment is clean; lays equipment out in convenient way; shows considerable skill and accuracy in making measures, transferring contents; well organised, plans ahead; works quickly; cleans up thoroughly on completion.*

*E fails to ensure equipment is clean; is disorganised in layout of equipment; is clumsy, careless and inaccurate when making measures, transferring contents; disorganised, no planning; works noisily, inefficiently; fails to clean up on completion.*

**Student self assessment**

Often it is as important for students to be able to assess the standard of their workmanship as it is to be able to produce a quality product.

When it is possible to provide a set of reliable criteria, the student can be required to assess his/her own product. In at least some cases the student assessments should be checked by the teacher. Care must be taken when using this system to ensure that the mark/re-mark agreement of student and teacher using the assessment scheme is high so the teacher is confident that the student can assess accurately.

## Project and field work assignments

Project and field work differ from practical work largely in terms of the environment in which they take place. They are often done outside the college and time allocations are usually generous, or should be. They also normally require a lot less teacher supervision.

*Examples of projects and field work*

- *Survey the neighbours in your street to find out their attitudes to smoking and alcohol.*
- *Measure the flow of traffic past your house between 7.00 p.m. and 8.00 p.m each day for a week and explain the reasons for any variations you observe.*

**Scope of activities**

A project is in effect a large-scale practical exercise from which time constraints have been largely removed. A very wide range of work is covered by the term 'project' and it can involve activities such as

- collecting and analysing data
- engaging in open-ended experiments in a laboratory
- making complete objects or parts of a larger project
- identifying and trying to solve problems.

**Choosing projects and field work assignments**

It is an advantage if students have some part in selecting their projects (perhaps from a list drawn up by the teacher). Because of the considerable effort that needs to be put into successfully completing this type of work, the element of student choice and hence student interest and satisfaction are important.

**Group work**

The various activities which make up project and field work assignments may be performed by individuals or by groups. Just which alternative is used needs to be taken into account when assessments are made.

**Interim reports**

Most project work tends to emphasise the finished product at the expense of the processes that have gone into achieving that product. Every effort should be made to assess the process skills even though this can be difficult, particularly if the work is being done in different places. One way to assess **process** is to call for interim progress reports, to initiate discussions with students, and to use a simple checklist to summarise how an individual is progressing.

**Role of teachers**

It is possible to take two rather extreme views of the teachers' role in project work. They can either provide the students with no help at all or alternatively treat the task as a largely teacher dominated exercise. Neither of these approaches is particularly productive and what is wanted is a balance between the two alternatives where the teacher acts as an important resource while encouraging students to work on their own as much as possible.

**Resources**

Good project work rarely occurs if the environment has inadequate resources. Therefore tasks set or chosen must be realistic—the teacher has an important role in deciding what is possible.

**Assessing for understanding**

In many projects it is inevitable that the students will have obtained help from a variety of sources, e.g. textbooks, parents, other students etc. What therefore becomes important in assessing project work is finding out whether the student **understands** what he or she has done. Oral assessments are a useful way of finding this out. (See page 74)

Exercise III(3) on page 80 provides an example of project assessment.

## Written assessments

The types of written assessment items used in TAFE range from the highly subjective to the highly objective—from items in which the students write an essay to those where they simply write a letter or tick a box.

The categories of written assessment items which will be considered in this section are:

Essay
Structured Essay
Short Answer
Multiple Choice
True-False
Matching Pairs

This is by no means the complete set of written items used in assessment; however, it does give the more commonly used ones. Readers interested in exploring other types should consult some of the texts on test construction listed in the references on page 77.

**Essay questions**

Essay questions are also known as 'extended response' and 'open ended' questions. The characteristics of the essay are its emphasis upon freedom of expression and creativity as well as upon depth and scope of knowledge. Hence, essay questions are usually used for measuring higher order cognitive abilities; for example, a student's ability to focus his/her knowledge on a problem, to relate and integrate relevant data, and to organise a coherent response.

The more creative forms of essay writing have limited applicability in many TAFE areas but teachers should not dismiss this form of assessment lightly. Creative writing can be a powerful source of stimulus for other classroom activities. A poetry writing assignment for the plumbing apprentices could brighten up an otherwise dull day although such an activity should not be attempted by the faint-hearted.

In essays with less emphasis on free expression the expectations are for an organised and reasoned argument on the set topic.

*Open-ended Essays*

*Example I:*

> *Describe the methods used in larger factories to correct metal distortion resulting from flame cutting.*

*Example II:*

> *Describe the design and operation of sprinkler systems installed to extinguish fires in office buildings. Include in your answer details of the systems triggered by melting solder and breaking glass.*

Essay questions that are not well constructed typically give the student very little indication of what is required in answering.. For example, the essay question, 'Discuss the safety requirements of workers in the metal trades industry', could be answered by writing a book. Questions like this create problems for both the student and the teacher. Both have to make judgments of what is an appropriate or satisfactory answer.

A suggested list of abilities that can be usefully assessed by essay questions include:

- explaining a cause and effect relationship
- describing applications of principles
- presenting relevant arguments
- formulating tenable hypotheses
- formulating valid conclusions
- stating necessary assumptions
- explaining methods and procedures
- describing the limitations of data
- producing, organising and expressing ideas
- integrating learning in different areas
- creating original forms (e.g. describing an experiment)
- evaluating the worth of ideas
- selecting important facts and ideas.

(From Gronlund, 1976)

**Guidelines for preparing essay questions**

If you choose to use an extended-response essay question you should be careful to ensure that:

i students are not being required to simply recall knowledge or information they have learnt. (An objective type of test will be a better way of doing this.)

ii students know what is expected of them. The choice of the lead word is particularly important: 'Describe' and 'Discuss' are very general words and usually allow for broad answers. 'Explain' is more specific, and 'Explain the relative contributions of . . . ' even more so.

iii in answering the question the student will be able to show his/her ability to describe, analyse, argue, synthesise and summarise. It is in assessing these abilities that an essay question is particularly appropriate.

iv if the essay is testing language skills, then some guidance should be given of the range of responses required. For example, the question should clearly indicate whether narrative, descriptive, argumentative, etc., answers are required.

v if aspects like spelling, punctuation, content, clarity, sequence, style, etc., are being marked, then this should be made clear to the student.

vi the student should know how much time there is to answer. Particular care must be taken to ensure that the question can be satisfactorily answered in the time available.

**Note:**

- Points iv, v and vi all relate to the DIRECTIONS FOR ANSWERING which should be a preamble to any test.

vii the item does **not** begin with the words 'What', 'Who', or 'List', as these do not usually require more than a short response.

viii unless what is being measured is an attitude or the defence of a position, do not ask: 'In your opinion' or 'What do you think'.

ix the question is 'tested' by writing an ideal answer to it. This allows the test constructor to check on the reasonableness of the question and the adequacy with which it is stated.

If possible, have someone else who should be able to answer the question try to answer it. Comparison of the two answers should show up any weaknesses in the question or suggest additional ways of improving it.

Criteria for marking should also be prepared at the time the questions are written. This means writing down what things will be rewarded and what things will be penalised. Different weightings are usually applied to different points made in an answer—the 'ideal answer' referred to in ix of the above guidelines is a good starting point for establishing the criteria for marking.

**Structured essay questions**

Structured essay questions are one way of removing any doubt the students may have about what is required of them. A good structured essay presents a well-defined task that improves the reliability of marking while at the same time allowing the students sufficient freedom to organise and express their thoughts. The form and scope of answers are limited because students are told specifically the context that their answers are to take.

*Structured Essays*

*Example I:*

*With the aid of a diagram, explain the operation of the soda-acid type of fire extinguisher. Label the diagram. Write the equation(s) showing the reaction that takes place when the extinguisher is put into operation.*

*Example II:*

*Four of the methods used to correct distortion of metals after flame cutting are:*

- *hydraulic presses*
- *beam benders*
- *stretching by hammer blows*
- *contra-heat.*

*Consider the use of these methods in typical jobs and explain their usefulness in terms of:*

- *the availability of equipment*
- *metal thickness*
- *likely damage to materials and*
- *the amount of time taken.*

**Guidelines for preparing structured essay questions**

If you choose to use a structured-response essay question, you should:

- avoid lengthy explanations of what is required; structured-response questions often become too wordy in an attempt to clarify what is wanted.

The points made under Guidelines i to ix should also be considered.

**Short answer questions**

Short answer questions are simple direct questions which can be answered by a number, a word, a short phrase or, at the most, a sentence or two.

Included in this group are sentence completion (or missing word) items and cloze items. The term 'cloze' comes from the idea of closure in psychology. Closure refers to the human tendency to fill in or complete something that appears to be incomplete. For example, 'By eliminating every fifth _____ in a sentence and _____ the students to write _____ the missing word, we _____ get an accurate idea _____ how well the passage _____ been understood'.

> (Answer: By eliminating every fifth *word* in a sentence and *asking* the students to write *in* the missing word, we *can* get an accurate idea *of* how well the passage *has* been understood.)

Completion tests can be particularly useful in measuring the student's retention of specific ideas or in assessing mathematical computational skills.

Short answer questions lie near the middle ground between the more objective and more subjective items. There should be few judgments to make about the answers to a short answer question. The marker's main problems should be confined to such things as deciphering the students' writing and deciding what to do about things like spelling and punctuation errors.

*Short Answer Questions*

*Example I*

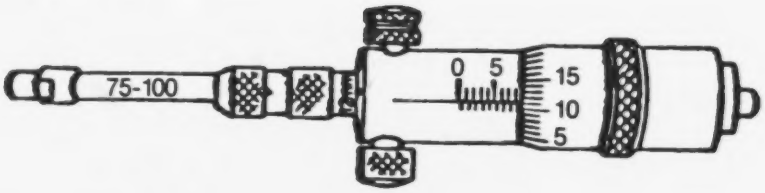
*Name the three basic operations performed on a lathe*

_____

_____

_____

*Example II*




*The reading on the inside micrometer shown is* _____

**Guidelines for writing short answer questions**

Things to remember when preparing Short Answer Test items are:

i to ensure there is a definite correct answer.

ii to be particularly careful in using sentences taken from textbooks. These often need their context if they are to make sense.

iii to be precise. 'Who was Bob Menzies?' is a vague question (and has a number of answers).

iv when some precision or degree of accuracy is required, as in science or maths calculations; to make it clear what level of precision is required (e.g. the number of significant figures). Similarly, if units of measurement are required, say so.

v when requiring completion of sentences; not to leave so many gaps that the sentences become puzzles.

*For example, a poor sentence completion exercise would be:*

*Welding is a* _____ *which involves* _____ _____ *with* _____.

**Multiple choice questions**

The most popular of the objective test items is the multiple choice question.

In these questions the students must choose one answer from among a given set of answers. Often there is **one definite answer**, but sometimes the task is to choose **the best answer**, or to choose the **one incorrect answer** among the choices. Once again it is important that the students know precisely what is expected of them.

Unfortunately, as we so often find in education, test developers have created their own special language to describe the questions. To understand this language, it is necessary to know the following definitions.

Stimulus Material:

Special information on which an item is based; this may be a paragraph, a diagram, a photograph, etc. Stimulus material is not always necessary.

Stem:

The part of the item which states the problem. Stems may be questions, directions, or incomplete statements.

Alternatives:

The choices from which the student is expected to select his/her answer. (Sometimes referred to as Options.)

Most multiple choice items usually offer 3 to 5 alternatives with 4 being the most common.

Key:

The correct alternative

Distracters:

The incorrect alternatives.

*Multiple-choice Question*

*The scientific name of the domestic cat is Felis domestica.* } *Stimulus*

*The word 'domestica' classifies the cat's* } *Stem*

| | | | |
|---|---|---|---|
| *A* | *family* | *Distracters* | *Alternatives* |
| *B* | *order* | | |
| *C* | *genus* | | |
| *D* | *species* | *Key* | |

**Guidelines for writing multiple choice questions**

The points to consider when constructing multiple-choice items are:

i present a simple, well defined problem in the stem.

*Example of a poorly-defined problem*

*Of all industrial accidents in Australia the most serious types are those involving*

*A muscle strains*

*B burns*

*C electrical shocks*

*D cutting equipment.*

The stem of this item has not defined the problem. We do not know the basis upon which to decide an answer. It could be which is most costly to the economy, most likely to cause death, most likely to cause permanent incapacity to individual workers, etc.

ii ensure that the alternative answers all relate to the stem in some way.

*Example of poorly-related alternatives*

*Which one of the following men invented the telephone?*

*A Bell* *C Pasteur*

*B Salk* *D Marconi.*

This question is concerned with an inventor in the field of communications; only Bell and Marconi were inventors in that field. Salk and Pasteur were biologists. Thus, two of the distracters offered do not relate to the stem.

iii include as much of the item as possible in the stem.

The stem should contain the central problem so that the students will have some idea as to what is expected of them, and some tentative answer in mind, before they begin to read the alternatives.

*Example of how to improve an item by transferring information into stem*

*Contrast the poorly constructed item:*

*Pine Knots*

*A should be covered with shellac before painting*

*B will stop giving off pitch if they are treated with raw linseed oil*

*C can sometimes be sealed with turpentine*

*D do not need any protective coating.*

*with the improved version:*

*In order to avoid paint discolouration, when painting pine boards containing many knots, which one of the following would you use to cover the knots?*

*A shellac* *C turpentine*

*B raw linseed oil* *D lacquer.*

iv the stems of the questions should be brief and to the point. (A mathematics item should not become a test in reading comprehension).

Illustrations are often useful in presenting the problem simply and clearly.

*Example of how illustration can improve a question*

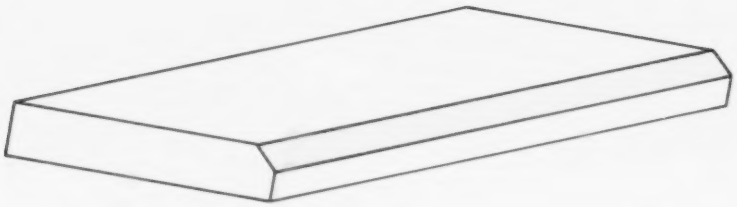
*Contrast:*

*A slanting out on a board that intersects two perpendicular surfaces, but not two parallel surfaces, is properly called*

*A a bevel* *C a slant*

*B an angle* *D a chamfer.*

*with the following:*



*In the drawing, the shaded part of the board is called*

*A* *a bevel* *C* *a slant*

*B* *an angle* *D* *a chamfer.*

v the alternatives must be grammatically consistent with the stem. Too often test writers provide clues to the answer by not attending to this requirement.

*Example of an item with bad grammar*

*Soda-ash fire extinguishers must NOT be used with an*

*A* *petrol fire* *C* *wood fire*

*B* *electric fire* *D* *plastics fire.*

The use of 'an' eliminates A, C, and D as possible answers.

vi make sure that the answer to any item cannot be found in other parts of the test. It is important to consider the test as a whole, because information supplied in one item can give away the answer to an item elsewhere in the same test.

vii do not include irrelevant or unnecessary information in the stem. This confuses the students as they must deal with the irrelevancies as well as the substance of the question and the result is a 'trick' question.

viii avoid using items with double negatives.

It is very confusing to be asked to realise that it is **not** true that Huon Pine is **not** grown in Tasmania.

ix avoid using items where the student must choose the incorrect alternative.

Sometimes it is much easier to think of three correct alternatives to a question than three incorrect ones. Indeed sometimes it seems to be impossible to think of three incorrect statements and hence the need to use this format occasionally.

However, including a large number of such items in a test can introduce an element of confusion as the students switch from negative to positive to negative and so on.

When writing negative items, always emphasise the negative and try to put it at the end of the stem.

*Example of using emphasis in questions asking for incorrect alternative*

*All of the following are desirable practices when preparing multiple choice items, EXCEPT*

*A stating the stem in a positive form*

*B using diagrams to cut down on the number of words*

*C emphasising any negative words that are used*

*D shortening the stem by lengthening the alternatives.*

x avoid mutually inclusive and mutually exclusive alternatives as much as possible. (i.e. 'all of these', 'none of A, B, or C', 'A and B only', etc should be used sparingly).

xi avoid mixed sets of alternatives which bear little relation to each other. These are really a set of true/false questions in disguise.

*Example of a multiple-choice question with an incorrect mix of alternatives*

*All of the following are true EXCEPT one. Which one?*

*A aluminium expands more than copper for the same increase in temperature*

*B a temperature of $0^{o}C$ is equivalent to a temperature of $273^{o}K$*

*C heat energy is used to produce electrical energy in a thermocouple*

*D water boils at a lower temperature on high mountains because of the humidity.*

This question covers a wide range of topics in 'physics', namely expansion of metals, temperature conversion, energy conversion and effects of air pressure. They bear little relation to each other.

(Note that each alternative would be suitable as a true-false item.)

xii try to have all alternatives approximately the same length or vary the length of the correct answer to eliminate length as a clue.

There is a tendency for the correct answer to be longer than the other alternatives because of the need to qualify statements to make them unequivocally correct. This, of course, provides a clue to the test-wise student.

*Example of item in which length of alternatives gives clue to answer*

*The term 'side-effect' of a drug refers to*

*A additional benefits from the drug*

*B the chain effect of drug action*

*C the influence of drugs on crime*

*D any action of a drug in the body other than the one the doctor wanted the drug to have.*

To overcome this problem it is often necessary to 'pad out' some of the incorrect alternatives and to deliberately write long incorrect alternatives in other questions.

xiii vary the position of the correct answer in a random manner.

The correct answer should appear in each alternative position (A, B, C, or D) approximately an equal number of times. Always check through a test before administering it to ensure that correct answers appear randomly.

xiv list the alternatives in some logical order, if this is applicable.

Numerical alternatives should be in either ascending or decending order.

*Examples of how to 'order' numerical alternatives*

*The product of 12 and 5 is*

| | | | |
|---|---|---|---|
| *A* | *7* | *C* | *60* |
| *B* | *17* | *D* | *125* |

*OR*

*The product of 12 and 5 is*

| | | | |
|---|---|---|---|
| *A* | *125* | *C* | *17* |
| *B* | *60* | *D* | *7.* |

**True-False Questions**

Basically, this type of item presents a question to which the student responds by choosing between two alternatives, of which only one is correct. The alternative answers are usually 'true' and 'false', but other forms which may be used are: yes-no; right-wrong; correct-incorrect.

*True-False Questions*

*Circle the correct answers:*

| | | | |
|---|---|---|---|
| *I* | *Air is a mixture of gases* | *T* | *F* |
| *II* | *Hydrogen sulfide is insoluble in water* | *T* | *F* |

**Guidelines for writing True-False questions**

When constructing True/False items be sure that you:

i keep the statements brief.

ii use statements where there is no exception to them being true or false. Avoid partly true or partly false items.

iii avoid including two or more ideas in one item.

iv avoid including trivial items. The true/false item should be used to encourage thought-provoking responses.

v avoid using words like 'all', 'always', 'no', 'never'. (The test-wise students realise these are usually false).

vi avoid broad general statements. Most broad generalisations are false unless qualified, and the use of qualifiers provides clues to the answer. Qualifiers such as 'usually' and 'sometimes' generally come with true statements.

vii avoid using statements lifted straight from textbooks.

Rephrase or modify the wording from a textbook so that you are measuring understanding rather than rote memory. Avoid wordings that are learned by rote, by presenting specific applications.

*Example of how to improve a rote-memory true-false question*

*For every action there is an equal and opposite reaction.*

*T F*

*is a poor item, but*

*When a hand pushes on a wall with a certain force, the wall pushes back on the hand with the same force.*

*T F*

*is acceptable.*

**Matching pairs questions**

These items usually consist of two lists in which the information is related. The student is told how the matching is to be done.

*Matching pairs question*

*Example:*

*Match all the fish dishes in List 1 with the appropriate sauces in List 2 by writing the number of the sauce in the box alongside the name of each fish dish.*

| | *List 1: Dishes* | | *List 2: Sauces* |
|---|---|---|---|
| ☐ | *sole colbert* | *(1)* | *beurre noisette* |
| ☐ | *sole goujon frit* | *(2)* | *tomato sauce* |
| ☐ | *plaice a l'Orly* | *(3)* | *beurre maitre d'hotel* |
| ☐ | *grilled herring* | *(4)* | *tartare sauce* |
| | | *(5)* | *mustard sauce* |

*[Answers from top to bottom are (3), (1), (2), (5)]*

The content of matching exercises can be related to such things as techniques or processes and their uses, parts and functions, rules and examples, terms and definitions; or at a more complex level, cause-and-effect relationships, and theoretical statements and experimental bases.

**Guidelines for writing matching pairs questions**

When writing matching-pair items be sure that you:

i put the question column in the left and the response (answer) column on the right.

ii give clear and precise directions (e.g., students need to be aware that they must match all the items in the left hand column, not just one pair).

iii keep the two lists of unequal size. This prevents students obtaining a correct answer by the process of elimination.

iv keep the number of items to be matched between about 4 and 10. (If there are more the exercise can become tedious as the lists get read over and over).

v ensure the elements in each list can be related in some way.

*Example of a poor matching pairs question*

*Match the terms in the left hand column with definitions in the right hand column by writing each of the numbers corresponding to the definitions in the appropriate boxes.*

| | |
|---|---|
| ☐ *nap* | *(1) used to finish hems* |
| ☐ *stay stitching* | *(2) supports garment shape* |
| ☐ *selvedge* | *(3) narrow edge of woven fabric* |
| ☐ *interfacing* | *(4) machine stitching within seam allowance* |
| | *(5) short fibres on fabric surface.* |

This is a poor example; the items in the question column bear little relation to one another; 'nap' and 'selvedge' are fabric characteristics; 'stay stitching' is a technique used in assembling a garment; and 'interfacing' is a type of fabric used in the construction of a garment. Hence, little or no discrimination between the items is required. Secondly, the one incorrect option in the response list is just a random choice and does not relate to the rest of the list except for being within the garment construction area.

## Oral and listening assessment

### Oral assessment

This type of assessment requires the student to make a spoken response to a given item. The teacher grades or marks the response using a checklist in much the same way as he or she would in a practical assessment.

Oral tests almost always require one-to-one teacher-student situations. They are demanding in terms of time and can be threatening, particularly for the shy or nervous student.

*Oral Questions*

*Ask the student to*

- *speak on a given topic for 5-10 minutes*
- *name a piece of equipment and explain what it is used for.*

When setting oral tests you should remember that:

i before giving an oral assessment the student should know which abilities are to be tested. You should also determine in advance the standards or levels of performance expected;

ii some abilities will be more important than others and you need to be sure about the weighting given to these abilities. For example, reading aloud from a book is a relatively simple task compared to giving a talk on a topic;

iii oral tests can be particularly useful as a support to practical, project or field assessments. By asking students to speak about their work you can find out how much was their own work and whether or not they understood what they were doing;

iv oral assessment takes a lot of time and is best used for those parts of a subject where only an oral test will do. For example, in language work only an oral test can test the students' ability to pronounce words correctly.

### Listening assessment

In this type of assessment the students have to listen to some material and are scored on their ability to comprehend or understand what they hear. Because students spend a very large amount of their time listening (both in college and away from it) developing their listening skills (or aural abilities) is most important. The ability to take shorthand is highly dependent upon listening skills.

*Listening Questions*

*Students are required to answer a number of questions after listening to*

- *a recording of a broadcast on some course-related topic*
- *a set of directions on fire drill.*

When preparing listening tests you need to remember that:

i preparation takes time. The use of a cassette recorder is recommended as a reusable collection of listening material can be prepared.

ii clarity of expression is important. The task of listening must not be complicated by poor quality presentation.

iii the listening test is not meant to test the students' ability to memorise what they hear. Therefore, in preparing the recording or reading to the students, the material will need to be repeated—possibly several times.

## Special purpose assessments

Most assessment in TAFE concerns knowledge, thinking skills, and manual skills.

In addition to the commonly used tests already described there are a number of others which could be useful. Whether teachers use any of these will really depend on what they want to test. They must decide, 'Is this the most appropriate test for what has to be tested'? There needs to be a good reason for using other testing methods. However, the following two are worth special consideration—other types can be found in the literature. (See References, page 77)

### Open book test

Here the students may use their books, but under examination conditions; that is, a time limit, no talking, etc. In some ways it is like a project or assignment with a more severe set of conditions.

This can be a very useful technique in TAFE subjects where students must work with manuals (e.g. roof construction manuals, laws and regulations). When using manuals in a typical on-the-job activity, the open book test is very relevant.

### Disclosed topics test

By announcing the actual questions to be given in a test or examination the teacher gives the students the opportunity to prepare answers and ensure a better chance of success.

This type of test can act as an incentive to student learning. One effective way of using a disclosed topics test is to ensure that the announced questions give a good coverage of the whole of the subject, but to only use a sample of these in the final test. Of say, 10 announced questions, the teacher may only use 5, the assumption being that the ability revealed in answering the sample would apply to the whole set. This type of test should not be used for objective items; it needs to stimulate the students to prepare quality answers, not simply to learn by rote.

**FINAL NOTE**

Many types of test types have been identified in the previous sections.

As a general rule, a combination of test types is better than using one type alone. However, this rule should not be interpreted too literally. A combination of test types in a single assessment session can be confusing to the student. The point is to use different methods over a period of time rather than all at the one time.

When students make mistakes because of the complexity of the test format the test is no longer testing the objectives it set out to test.

Even in more extended testing sessions such as end-of-term or end-of-year examinations, when it is necessary to use more than one method, it is important to keep each method separate. For example, there should be a section for multiple choice questions, another for short answers, and so on. The question types should not be changing randomly from one to another throughout the test.

## REFERENCES


Bloom, B. S., Hastings, T. T., & Madaus, G.F. Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971.

Brown, F. G. Measuring classroom achievement. New York: Holt, Rinehart and Winston, 1981.

Ebel, R. L. Essentials of educational measurement. Englewood Cliffs: Prentice Hall, 1972.

Ebel, R. L. Measuring educational achievement. Englewood Cliffs: Prentice Hall. 1976.

Gronlund, N. E. Preparing criterion-referenced tests for classroom instruction. New York: MacMillan. 1973.

Gronlund, N. E. Measurement and evaluation in teaching. New York: MacMillan. 1976.

Popham, W. T. Criterion-referenced measurement: An introduction. Englewood Cliffs: Educational Technology Publications. 1971.

Rowntree, T. Assessing students. London: Harper and Row. 1971.

Swezey, R. W. Individual performance assessment: An approach to criterion-referenced test development. Virginia: Reston. 1981.

Thorndike, R. L. (Ed.). Educational measurement (2nd ed.). Washington: American Council on Education. 1971.

## EXERCISE III (1)

## A COMPARISON OF MULTIPLE-CHOICE AND ESSAY TEST ITEMS

Below is an incomplete table which compares multiple-choice and essay test items.

Using the spaces provided, complete the table of comparisons showing the advantages/disadvantages of each type. You can check your answers by turning to page 86.

| | MULTIPLE-CHOICE | ESSAY |
|---|---|---|
| Learning Outcomes Measured | Good for measuring knowledge, comprehension and application outcomes; generally inadequate for analysis, synthesis and evaluation outcomes. | |
| Sampling of content | | The use of a relatively small number of items results in limited coverage, which makes representative sampling of content difficult. |
| Preparation of Items | Preparation of good items is difficult and time-consuming. | |
| Scoring | | Subjective, difficult, and less reliable. |
| Factors that Distort Student's Scores | Reading ability and guessing. | |
| Probable Effect on Learning | | Encourages students to organise, analyse, synthesise and express their own ideas. |

## EXERCISE III (2)

## CHOOSING THE MOST APPROPRIATE TEST ITEM

Below are eight objectives from a list of objectives in a basic life skills course.

Select the test types which would be the most appropriate to use in assessing each objective.

Suggested answers are on page 87.

| OBJECTIVE | MOST APPROPRIATE TEST TYPES |
|---|---|
| Recognise 1/2 and 1/4 | |
| Find the area of a rectangle in square metres | |
| Read bar graphs | |
| Recognise words that rhyme | |
| Design house that incorporates energy-saving features | |
| Identify the general mood of a story | |
| Participate in group games that call for strict observance of rules | |
| Write a simple greeting for a forthcoming birthday of a relative. | |

## EXERCISE III (3)

## PROJECT ASSESSMENT

Suppose the college was asked to provide a set of brick barbecues on a vacant block of land alongside some retirement cottages.

Five barbecue sites are to be built, and as teacher-in-charge you have five groups of students who are each going to build one barbecue.

You have also decided to assess the work of each group by rewarding the following aspects of the task:

- Planning and Design
- Originality
- Quality of workmanship
- Awareness of links between task and life in the community.

Indicate how you would grade each of these aspects

1. Planning and Design _____

_____

2. Originality _____

_____

3. Quality of workmanship _____

_____

4. Awareness of links between task and life in the community _____

_____

_____

What technique would you use to assess No. 4, the 'Awareness' aspect?

_____

_____

_____

_____

(Suggested answers are on page 88).

## EXERCISE III (4)

## ORAL ASSESSMENT

Consider the oral item:

> Briefly explain the safety rules that apply in the workshop/laboratory.
>
> (This exercise should take the student about two minutes)

The talk is to be judged in term of its

> INTELLIGIBILITY
>
> ACCURACY
>
> VOCABULARY

Plan a five point checklist for each of these aspects so that you can grade the student on each one. Indicate clearly what it is that you will be listening for when making your judgments.

(Suggested answers are on page 89).

## EXERCISE III (5)

### ITEM REVIEW

Review the items below and attempt to relate any mistakes you find to the suggestions on writing multiple choice items found on pages 66 to 71.

Use the space provided to explain what, if anything, is wrong with the item and what could be done to improve it.

(Note: It is not always necessary to know the correct answer to a question to carry out a review.)

1. Steam is

   A given off when carbon burns

   B produced by the burning of natural gas

   C a by-product of sulfur burning

   D a by-product of magnesium burning

   REVIEWER'S COMMENTS

   _____

   _____

   _____

   _____

2. When written as a percentage, 7/20 equals

   A a number less than 28%

   B exactly 28%

   C a number greater than 28%

   D 35%

   REVIEWER'S COMMENTS

   _____

   _____

   _____

   _____

3. Which one of the following games is the odd one out?

A billiards C badminton

B baseball D soccer

REVIEWER'S COMMENTS

4. The discoverer of the Law of Gravity was

A Milton C Ronald Reagan

B Newton D Elizabeth Taylor

REVIEWER'S COMMENTS

5. A boy has three 5 cent coins, three 10 cent coins and one 20 cent coin. How much money does he have?

A more than 40 cents C more than 60 cents

B more than 50 cents D more than 70 cents

REVIEWER'S COMMENTS

6. The position of Poet Laureate offers a modest salary for relatively little work.

The laureateship is a

A imposition
B sinecure
C embarrassment
D emolument

REVIEWER'S COMMENTS

7. In a class of forty children, one fifth come by bus and three fifths walk. How many come by bike?

A 8
B 16
C 32
D None of the above

REVIEWER'S COMMENTS

8. All of the following are true EXCEPT one. Which one?

A The acidity of the human stomach is due to the presence of hydrochloric acid.

B Bile is made in the pancreas.

C Incisor teeth are used for cutting food.

D The pancreatic duct conveys food digesting enzymes to the small intestine.

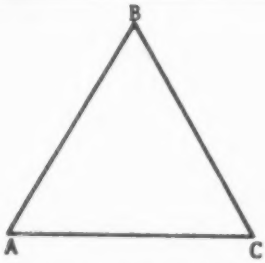
REVIEWER'S COMMENTS

9. Which of the following compounds is a gas?

A water C carbon dioxide

B ethanol D silicon dioxide

REVIEWER'S COMMENTS




10. In the figure ABC, if AB = BC, than ABC is an

A right angle triangle C equilateral triangle

B scalene triangle D isosceles triangle

REVIEWER'S COMMENTS

(Suggested answers are on pages 90 and 91).

## SUGGESTED ANSWERS—EXERCISE III (1) [Pg 78]

| | MULTIPLE-CHOICE | ESSAY |
|---|---|---|
| Learning Outcomes Measured | Good for measuring knowledge, comprehension and application outcomes; generally inadequate for analysis, synthesis and evaluation outcomes. | Inefficient for knowledge outcomes; good for comprehension and application outcomes; best for analysis, synthesis and evaluation outcomes. |
| Sampling of content | The use of a large number of items results in broad coverage, which makes representative sampling of content feasible. | The use of a relatively small number of items results in limited coverage, which makes representative sampling of content difficult. |
| Preparation of Items | Preparation of good items is difficult and time-consuming. | Preparation of good items require time and thought, but easier than the preparation of objective items. |
| Scoring | Objective, simple, and more reliable. | Subjective, difficult, and less reliable. |
| Factors that Distort Student's Scores | Reading ability and guessing. | Writing ability and bluffing. |
| Probable Effect on Learning | Encourages students to memorise and understand the ideas of others. | Encourages students to organise, analyse, synthesise and express their own ideas. |

## SUGGESTED ANSWERS—EXERCISE III (2) [Pg 79]

| OBJECTIVE | MOST APPROPRIATE TEST TYPES |
|---|---|
| Recognise 1/2 and 1/4 | Multiple-Choice |
| Find the area of a rectangle in square metres | Multiple-Choice, Short-Answer |
| Read bar graphs | Multiple-Choice, Short-Answer |
| Recognise words that rhyme | Aural, Multiple-choice |
| Design house that incorporates energy-saving features | Project |
| Identify the general mood of a story | Structured-Response, Short-Answer, Multiple-Choice |
| Participate in group games that call for strict observance of rules | Practical (checklist) |
| Write a simple greeting for a forthcoming birthday of a relative. | Structured-response essay, short-answer. |

## SUGGESTED ANSWERS—EXERCISE III (3) [Pg 80]

Each aspect could be graded on a 1 to 5 scale as follows

| | |
|---|---|
| Excellent | 5 |
| Good | 4 |
| Satisfactory | 3 |
| Fair | 2 |
| Poor | 1 |

Assessing 'awareness' is probably best done by oral questioning. This can be done both while the project is underway and when it is completed.

Possible questions relevant to 'awareness' are:

- Which groups do you think are going to use this barbecue?
- Why did you choose this particular design?
- Where are there other barbecues in the community?
- What has been the reaction of people in the retirement cottages to this project?

After questioning the students, the teacher should complete an 'awareness checklist' using a 1 to 5 scale as above.

This should be done on more than one occasion.

## SUGGESTED ANSWERS—EXERCISE III (4) [Pg 81]

Three possible checklists are presented below.

**INTELLIGIBILITY**

4 = Perfectly or nearly perfectly intelligible, no difficulty in following what student says

3 = Very clear but one or two lapses

2 = Generally intelligible but some words cannot be understood or are slurred

1 = Difficult to follow what is being said, frequent slurring of words

0 = Almost incomprehensible

**ACCURACY**

4 = Includes all the rules and stresses more important ones

3 = Includes all the rules but doesn't convey level of importance

2 = Knows most but not all rules

1 = Knows only one or two rules and then imperfectly

0 = Has only vague idea of rules, makes up inappropriate rules, shows disregard for safety

**VOCABULARY**

4 = Words used are suitable, knows correct terms for various safety equipment, none mispronounced, no inappropriate jargon

3 = Words suitable, knows most of terms for equipment, one or two words mispronounced or used inappropriately

2 = Words generally suitable, does not know terms for most of equipment, a few mispronounced or used inappropriately

1 = Many unsuitable words used, many mispronounced

0 = Vocabularly extremely limited and the few technical words used are generally mispronounced.

## SUGGESTED ANSWERS—EXERCISE III (5) [Pg 82]

1. Too much detail in alternatives that should be in stem.

   REWRITE

   Steam is given off during the burning of

   A carbon
   B natural gas
   C sulfur
   D magnesium.

2. Two alternatives are correct (C & D). Without knowing the objective that is being tested it is difficult to be sure what to do with this item. Assuming it is simply to accurately perform the calculation 7/20 x 100, the REWRITE could be

   When written as a percentage 7/20 equals

   A 7%/20
   B 7%
   C 20%
   D 35%.

3. The question does not give the criteria by which an answer can be arrived at. All could be correct. For example billiards is the only one played with more than one ball; baseball is the only one in which a player is rewarded for hitting a ball out of the playing area; badminton the only one in which racquets are used; and soccer the only one where the feet and head are used to propel the ball.

4. Only one of the four alternatives has anything to do with physics so the answer can be obtained without knowing that Newton actually did discover the Law of Gravity.

5. The actual amount of money is 65 cents; therefore, C is correct, but so too are B and A as the question stands.

   This could be rewritten to ask for the 'best' answer, or to give actual amounts in the alternatives.

6. The answer is obtained from the grammatical construct.

   Alternatives should be rewritten as

   A an imposition
   B a sinecure
   C an embarrassment
   D an emolument.

7. The inclusion of D confuses the issue. Some students might think the question contained a trick as the remainder did not necessarily have to come by bike, e.g., they might have come by car.

8. The statements are not closely related. These 4 alternatives would be better presented as four separate TRUE/FALSE questions.

9. Question fails to adequately define the problem. We need to know the conditions under which we will make our judgment because the temperature and pressure determine whether a substance is a solid, a liquid or a gas.

   REWRITE

   Which of the following compounds is a gas at normal room temperature and pressure?

   | | | | |
   |---|---|---|---|
   | A | water | C | carbon dioxide |
   | B | ethanol | D | silicon dioxide. |

10. Alternatives are not grammatically consistent with the stem. Diagram could be improved by making it look less like an equilateral triangle.

    REWRITE

    In the figure ABC, if AB=BC, then ABC is

    | | | | |
    |---|---|---|---|
    | A | a right angle triangle | C | an equilateral triangle |
    | B | a scalene triangle | D | an isosceles triangle. |

# CHAPTER IV

# TEST SPECIFICATIONS

## WHAT THIS CHAPTER IS ABOUT

This chapter is concerned with the development of test specifications by the classroom teacher. Although not directed specifically at those involved in TAFE Authority-wide testing the principles of test specification development are very similar in both cases.

The chapter looks at two aspects of classroom testing, namely:

- overall planning for the assessment of the student learning
- planning for individual tests.

*OBJECTIVES*

*After completing this chapter, you should be able to:*

- *distinguish between planning an assessment program for a whole subject and planning for a single test*
- *define what is meant by a table of specifications and recognise alternative approaches exist*
- *explain the importance of developing a table of specifications for classroom assessment*
- *name the steps in developing a table of specifications*
- *develop a table of specifications using a two-dimensional grid with relative weights assigned to every cell of the grid.*

## INTRODUCTION

This chapter is concerned with test specifications of relevance to the classroom teacher. The way a teacher goes about developing specifications depends on whether the subject is examined externally or internally.

Teachers of subjects which are externally examined need to relate their testing to the requirements of external examiners as set down in the syllabus.

When subjects are examined internally the teachers have much more control over decisions about what content and abilities should be tested.

However, there also many subjects developed by teachers within colleges that lack syllabus statements outlining these college-based subjects with sufficient precision for the person constructing a test. The need for precision when detailing the objectives of a subject is discussed in Chapter II. The need is further reinforced in this chapter as it will be shown that all tests must be related to learning outcomes that are stipulated by the objectives in the syllabus statement.

> *A traditional difficulty of students has been to understand what examiners and/or teachers expect of them when it comes to assessment.*

Students need to know how much of what has been taught will be tested, and what level of performance is expected. In the past many students would overcome this difficulty by studying old tests and old examination papers to gain the required understanding. However, this practice should be unnecessary because properly developed test specifications satisfying the students' need to know what it is they must be able to do after a period of instruction.

## PLANNING TESTS FOR A WHOLE SUBJECT AND A UNIT OF THAT SUBJECT

Generally, the testing of a whole subject is done by using a number of sub-tests each directed at a particular unit of the work within the subject. In such cases it becomes important to ensure that an overall assessment plan exists otherwise certain objectives can receive inadequate treatment. The inadequate treatment or neglect of some objectives stems from the basic problems that classroom teachers have with test construction.

### Achieving balance in a test

Many teachers have a tendency to overload their tests with items measuring knowledge of isolated facts and to neglect the more complex learning outcomes.

> *In the general studies area, for example, it is not uncommon to include a disproportionately large number of items which measure knowledge of names, dates, places, and the like. In engineering, the defining of terms and the naming of component parts and their uses are commonly overemphasised. In mathematics, computational skill is frequently the only learning outcome measured.*

In most syllabus statements the less complex learning outcomes are seen as less important than understandings, applications, interpretations, and various other 'thinking skills'. However they often receive undue prominence because the teacher finds it easier to construct test items which require the recall of this type of information.

Without a carefully developed test plan, ease of construction all too frequently becomes the dominant factor in selecting items and constructing tests. As a consequence, the tests measure a limited and biased sample of student ability and neglect many of the learning outcomes that are considered to be more important.

When the tests set during the course have such problems the cumulative effect is a poor testing program.

## Need for an overall test plan

Teachers need to begin their teaching program with plans to assess **all** the objectives of the course.

Including all instructional objectives in such a plan provides a clear picture of what is being measured by classroom tests. This clarifies the important role of testing in the total teaching process and at the same time should prevent an over emphasis on testing procedures. Each assessment procedure is seen in its proper perspective and its role in the total program is readily perceived. Constructing tests to fit the course objectives and the particular group of students being taught are important elements in this overall planning.

## TEST SPECIFICATIONS FOR A SUBJECT

The first step in the assessment of student learning of any teaching program is the setting down of test specifications. This involves making important decisions concerning:

- the content is to be tested
- the objectives (cognitive, affective and psychomotor) to be assessed
- the weighting is to be assigned to the various elements of content and objectives.

A convenient way of arriving at such decisions is to set up a table of specifications in the form of a test grid.

The following table of specifications has been developed for the subject 'Individual Development' in a Typing and Communication course in NSW.

**Table 1a**

**TABLE OF SPECIFICATIONS FOR THE SUBJECT: INDIVIDUAL DEVELOPMENT**

**(From Typing and Communication—NSW)**

| Objectives / Content Topics | Cognitive Domain | | | Affective Domain | | | Total Emphasis (percentage) |
|---|---|---|---|---|---|---|---|
| | Knowledge | Comprehension | Application | Responding | Valuing | Organising | |
| Exercise and Relaxation | 6 | 6 | 12 | - | - | - | 24 |
| Diet and Nutrition | 6 | 6 | 12 | - | - | - | 24 |
| Recreation and Leisure | 6 | 6 | 10 | - | - | - | 22 |
| Self-Perception | - | - | - | 9 | 4 | 2 | 15 |
| Inter-Personal Relationships | - | - | - | 9 | 4 | 2 | 15 |
| Total Emphasis (percentage) | 18 | 18 | 34 | 18 | 8 | 4 | 100 |
| | 70% | | | 30% | | | |

In assigning weights to the various elements of content and objectives a common procedure is to weight each area in terms of the time (number of weeks) to be devoted to it. The weights assigned to each objective should also reflect the importance the teacher attaches to that objective and the emphasis to be given to it in his or her teaching. (The emphasis will normally be specified or at least indicated in the syllabus document). Thus, if the syllabus stresses the comprehension and application of principles and generalisations, the teacher should give these objectives correspondingly greater weight.

> *The areas of content and the objectives in the table of specifications, should be assigned weights which reflect the emphasis given to them in the subject. When a lot of time has been given to teaching a particular topic, then that topic needs to be thoroughly tested and its assessment given a greater weight than is given to other topics.*

This will make it possible to construct tests that measure both content and objectives in a representative manner.

The assigning of numbers as shown in Table 1a gives the impression of a highly objective operation. This is basically a false impression as the assignment of weightings described above has a considerable degree of subjectivity attached to it. Indeed, not only must opinions about degrees of importance of topics be taken into account, but there is also the question of availability of test items.

A major problem for test constructors is the fact that questions on some topics are more difficult to write than others. There is no point in using a poor question in a test simply because it is on a topic that is not adequately covered. A good question must be found or the inadequate coverage of the topic accepted as one of the many unfortunate 'facts of life' in the art of test construction.

In practice, therefore, Table 1a would have evolved from a more pragmatic table of specifications that looked something like Table 1b.

**Table 1b**

**TABLE OF SPECIFICATIONS FOR THE SUBJECT: INDIVIDUAL DEVELOPMENT**

| Objectives / Content Topics | Cognitive Domain | | | Affective Domain | | | Total Emphasis (percentage) |
|---|---|---|---|---|---|---|---|
| | Knowledge | Comprehension | Application | Responding | Valuing | Organising | |
| Exercise and Relaxation | ✓ | // | /// | | | | |
| Diet and Nutrition | ✓ | // | /// | | | | |
| Recreation and Leisure | ✓ | // | /// | | | | |
| Self-Perception | | | | // | ✓ | | |
| Inter-Personal Relationships | | | | // | ✓ | ✓ | |
| | About 2/3 rds | | | About 1/3 rd | | | |

On this table the ticks show the probable initial thinking of the test constructor. This would have become more 'refined' as actual items were considered and with some juggling of numbers the final result was Table 1a.

In the classroom situation, with limited time and resources, a Table of Specifications of the type shown in Table 1b is quite adequate for most purposes.

## Comprehensiveness of subject test specifications

Subject test specifications must be comprehensive. Teachers need to include objectives other than those measured in written tests. In some subjects, practical performance is all important.

The table then becomes a general plan which includes the specifications of the overall testing program as well as suggestions concerning methods of assessment to be used in particular areas.

## Modular approach

TAFE courses are increasingly being presented in a modular form. This involves breaking the syllabus up into small units which are taught and assessed independently.

*Consider the following part of a module from an earthmoving operators course.*

*MODULE: BULLDOZERS*

*TOPIC: RIPPING TECHNIQUES*

*PERFORMANCE:* *Trainees will safely and efficiently use the rippers on all the available bulldozers.*

*CONDITIONS:*
- *Practical instruction.*
- *At least two different makes of bulldozer with rippers.*
- *Suitable practical work area.*

*STANDARDS:*
- *To instructor's satisfaction during practice.*
- *Satisfactory score on checklist.*

*ASSESSMENT:*
- *By instructor evaluation of practice performance.*
- *By checklist items as follows.*

*CHECKLIST FOR RIPPING TECHNIQUES*

| *ITEM* | *✓or X* | *COMMENT* |
|---|---|---|
| *RIPPING* | | |
| *1. Lowers rippers while moving.* | | |
| *2. Uses correct number of rippers for ground hardness.* | | |
| *3. Rips in a straight line.* | | |
| *4. Rips with the grain of the ground.* | | |
| *5. Adjusts ripper length to suit ground.* | | |
| *6. Does not turn while ripper is penetrating.* | | |
| *7. Retracts ripper while moving.* | | |

*INSTRUCTOR SIGNATURE:* _____ *DATE:* _____

This sub-module on ripping techniques is one of a set which makes up the Bulldozer Module of an Earthmoving Operators Course. By combining the performance objective, conditions and standards with the actual assessment instrument a mini test specification is produced. The whole module is assessed via the combination of the individual assessments on each sub-module.

This approach is an alternative to a Table of Specifications as set out in Table 1a.

**TEST SPECIFICATIONS FOR A UNIT OF A SUBJECT**

Once an overall test specification has been decided upon for a subject it is a relatively straightforward task to develop specifications for testing parts of that subject.

Test specifications for units are done in the same way as those for subjects except that a more detailed breakdown of both content and objectives is usually appropriate. Consider the following table for a unit on camera techniques.

Table 2

TABLE OF SPECIFICATIONS FOR A

CURRICULUM UNIT: BASIC CAMERA TECHNIQUES

| Objectives / Content Topics | Cognitive Domain | | | | Performance Domain | Total Emphasis (percentage) |
|---|---|---|---|---|---|---|
| | Knowledge | Comprehension | Application | Evaluation | Practical Skills | |
| Camera Types | 4 | - | - | - | - | 4 |
| Image Formation | 6 | 3 | 6 | - | - | 15 |
| Photographic Routine | 4 | 1 | - | - | 4 | 9 |
| Standardisation | 4 | 1 | - | - | 4 | 9 |
| Line Copy and Negative Assessment | 2 | 1 | - | 1 | - | 4 |
| Exposure Control | 4 | 1 | - | - | - | 5 |
| Fixed Aperture System | 5 | - | 5 | - | 8 | 18 |
| Fixed Ratio System | 5 | - | 5 | - | 8 | 18 |
| Development | 2 | 1 | - | 2 | 4 | 9 |
| Emulsion Characteristics | 2 | 1 | - | 2 | 4 | 9 |
| Total Emphasis (percentage) | 38 | 9 | 16 | 5 | 32 | 100 |

Table 2 uses the approach of breaking down the content into more precise elements.

An alternative approach is to break down the objectives into more specific elements as shown in Table 3 which sets out the specifications for a unit on weather in the subject Navigation.

Teachers should choose the technique which is most appropriate for their subject. It is important to ensure that a sufficient level of detail exists to make it clear what will be assessed.

The table of specifications for the subject should be made available to the students to help them plan their work. Whether or not tables of specifications for unit tests are given to the students depends to some degree on the subject being tested. Obviously for most practical tests they should be, but in some areas of theory this action may be inappropriate.

**Table 3**

**TABLE OF SPECIFICATIONS FOR A UNIT ON WEATHER**

| Content | OBJECTIVES | | | | | | |
|---|---|---|---|---|---|---|---|
| | Knows | | Understands influence of each factor on weather formation | Interprets weather maps | Performance Skill in | | |
| | symbols and terms | specific facts | | | use of measuring devices | constructing weather maps | |
| Air pressure | 2 (6%) | 2 (4%) | 3 (6%) | 2 (4%) | Observe pupils using measuring devices (Rating Scale 0-3 for each) | Evaluate maps constructed by pupils (Rating Scale 0-2 for each) | |
| Wind | 2 (4%) | 2 (4%) | 5 (10%) | 2 (4%) | | | |
| Temperature | 2 (4%) | 2 (4%) | 2 (4%) | 2 (2%) | | | |
| Humidity and precipitation | 2 (4%) | 1 (2%) | 2 (4%) | 2 (4%) | | | |
| Clouds | 2 (4%) | 2 (4%) | 1 (2%) | | | | |
| Total number of items | 10 | 9 | 13 | 8 | 4 | 4 | 48 |
| Percent of marks | 22 | 18 | 26 | 14 | 12 | 8 | 100% |

**SUMMARY**

Tables of specification should be drawn up for both overall subject test programs and single tests.

They are the blueprints which teachers use to build their tests.

There are three simple steps in the construction of a table of specifications.

i **Down the left hand side (vertical axis) of the table list the subject matter divisions (content topics) to be assessed**

More detail may be added as required.

ii **Along the top (horizontal axis) of the table list the learning objectives to be assessed**

These should be grouped into the major objective domains or categories and it is usual to use the more general learning objectives for these headings. More specific objectives may be indicated for each outcome if this detail is required.

iii **Allocate a weighting to each subject matter and learning objectives category**

These weightings, written as percentages, indicate the relative emphasis to be given to each topic and objective and, consequently, indicate also what proportion of the total mark allocation is to be devoted to each category.

In assigning percentages it is wise to be flexible. In practice a weighting should be something to aim at rather than be treated as a mandatory figure. A little juggling of the percentage figures at the end is perfectly acceptable and is often necessary.

Some curricula have been developed in ways which incorporate the test specifications into the curriculum documents as shown on pages 98 and 99. Provided adequate guidance is given on the weighting to be applied to the various test items this is an equally good way of planning an assessment program.

## EXERCISE IV (1)

## PREPARING APPROPRIATE TEST SPECIFICATIONS FOR A SECTION OF WORK

The set of objectives in Reading Comprehension given below relate to part of a bridging course in basic literacy.

Complete the test specification grid by showing how you would assess the attainment of these objectives. (Assume the teaching of these objectives has taken two months to complete and the test time available is 3 hours. The 3 hours can be spread over more than one day).

Indicate against each objective the type of item or items you would use, the number of items that would be required, and the percentage of marks you would give to that particular objective. (The total percentage must add to 100).

### TEST SPECIFICATION GRID FOR READING COMPREHENSION TEST (BASIC LITERACY)

| NO. | OBJECTIVE | NO. OF ITEMS | BRIEF DESCRIPTION OF ITEM TYPE | % OF TOTAL |
|---|---|---|---|---|
| 1 | Identify key sentence | | | |
| 2 | Restate main idea | | | |
| 3 | Give title to paragraph | | | |
| 4 | Give title to poem | | | |
| 5 | Identify general mood of story | | | |
| 6 | Identify setting of story | | | |
| 7 | Identify main character | | | |
| 8 | Demonstrate comprehension of passage by answering 'how' and 'why' items | | | |

| NO. | OBJECTIVE | NO. OF ITEMS | BRIEF DESCRIPTION OF ITEM TYPE | % OF TOTAL |
|---|---|---|---|---|
| 9 | Classify ideas under headings | | | |
| 10 | Sequence events as they happen in a story | | | |
| 11 | Perceive cause-effect | | | |
| 12 | Predict outcomes | | | |
| 13 | Infer traits of character in a story | | | |
| 14 | Distinguish between reality and fantasy | | | |
| | | | TOTAL | 100 |

(Suggested answers are on pages 106 and 107).

## EXERCISE IV (2)

Develop a table of specifications for a unit in a subject you are teaching or have taught. Include both content and objectives dimensions and assign relative weights to every cell.

NOTE: There is no 'Suggested Answer' for this exercise.

In a workshop situation the individual tables would be the basis for group discussion.

## SUGGESTED ANSWERS—EXERCISE IV (1)

| NO. | OBJECTIVE | NO. OF ITEMS | BRIEF DESCRIPTION OF ITEM TYPE | % OF TOTAL |
|---|---|---|---|---|
| 1 | Identify key sentence | 4 | Multiple-choice with items based on 4 short paragraphs | 4 |
| 2 | Restate main idea | 1 | Short-answer with suitable paragraph for which main idea must be restated | 4 |
| 3 | Give title to paragraph | 1 | Short-answer requiring title for suitable paragraph | 5 |
| 4 | Give title to poem | 1 | - choice of poem is important<br>- note the judging of this answer has a subjective element | 5 |
| 5 | Identify general mood of story | 1 | Structured-response requiring student to write on mood of story that has been studied | 10 |
| 6 | Identify setting of story | 4 | Multiple-choice questions of previously studied stories requiring identification of settings (4) and main character (4) | 4 |
| 7 | Identify main character | 4 | | 4 |
| 8 | Demonstrate comprehension of passage by answering 'how' and 'why' items | 20 | Multiple-choice questions based on passages provided as stimulus material | 20 |
| 9 | Classify ideas under headings | 5 | Matching-pairs—give students a number of headings and ask for ideas to be matched | 5 |
| 10 | Sequence events as they happen in a story | 1 | Short-answer—use known story, provide incorrect sequence and ask for correction (alternatively, use multiple choice question) | 5 |
| 11 | Perceive cause-effect | 1 | Essay—use story that has been studied and ask for reason why certain events occurred | 10 |

(cont) **SUGGESTED ANSWERS—EXERCISE IV (1)**

| NO. | OBJECTIVE | NO. OF ITEMS | BRIEF DESCRIPTION OF ITEM TYPE | % OF TOTAL |
|---|---|---|---|---|
| 12 | Predict outcomes | 1 | Essay—use unseen paragraph which sets up a problem situation and ask for essay on outcome | 10 |
| 13 | Infer personality of character in a story | 1 | Structured-response—require students to describe personality | 10 |
| 14 | Distinguish between reality and fantasy | 4 | True/False asking for reality/fantasy of previously studied situations | 4 |
| | | | TOTAL | 100 |

NOTE: With only three hours for testing, oral testing is not possible.

## CHAPTER V

## ITEM ANALYSIS

### WHAT THIS CHAPTER IS ABOUT

This chapter outlines the methods for analysing the results of classroom tests.

Such analysis enables a check to be made that the questions are effectively testing what the teachers set out to test.

*OBJECTIVES*

*After completing this chapter you should be able to:*

- *know the different techniques for analysing norm-referenced and criterion-referenced tests*
- *explain the terms item difficulty, discrimination and distracter analysis*
- *describe a suitable method for keeping statistics on teacher-made tests.*

### STATISTICS, VALIDITY AND TEST TYPES

Test statistics are important to the teacher as they provide objective information on how much is being learnt and how well subject matter is being taught. However, test statistics are only as good as the quality of the test items allows. The word that test constructors use to describe this quality is 'validity'.

> *Although there are different types of 'validity', the most relevant for typical classroom, laboratory, and workshop situations is content validity—that is, the extent to which a test item measures performance in the subject matter of the course that is being taught.*

To determine whether items are valid for a particular group, the educational objective being tested by each item should be examined. Teachers must make their own judgments on the relevance of items for their students.

In this chapter the issues of statistics, validity, and test type are related to multiple choice tests, however the principles can be applied to any type of test.

There are important differences in the way statistics are handled in norm-referenced and criterion-referenced tests. The issues are considered in the next two sections—Chapter I explained the difference between the test types.

## Statistics for norm-referenced multiple-choice tests

The validity of a test may be affected by items which are either too difficult or too easy for the students being tested; by items which do not discriminate effectively between more able and less able students; and by items which have unsuitable distracters.

Three pieces of statistical analysis which all teachers can perform on their own tests are:

- a determination of the **difficulty** of each item;
- an estimate of the **discrimination** of each item; and
- a **distracter analysis.**

These three procedures are described in the following sections.

### Difficulty

The difficulty of a multiple-choice item is given by the following formula:

$$\% \text{ difficulty} = \frac{\text{number of incorrect responses}}{\text{total number of responses}} \times 100$$

Each item is scored 1 for a correct response and 0 for an incorrect response. For summative tests such as end-of year examinations, items which all students answer correctly (i.e. 0% difficulty) or items which no student answers correctly (i.e. 100% difficulty) contribute nothing to separating the students. However, for diagnostic testing purposes, such items may indeed be valuable.

Test constructors in this country also use the term Item Facility as an alternative to Item Difficulty. The facility of an item is the percentage of students giving correct answers.

Teachers need only use one of these terms and item difficulty, expressed as a percentage, will be the one used in this chapter.

*The higher the percentage difficulty the harder the item. If the difficulty, d = 80%, then 8 out of every 10 students are getting the item wrong.*

Teachers should, however, be warned to check the way the term is being used when they are reading about item difficulties. American literature uses the term in exactly the opposite way to us. (See Theobald, 1974; Izard, 1977.) As if to demonstrate the perverseness of the test statistician, the US literature defines difficulty in terms of the percentage answering correctly. They are, in fact, using the term in the way item facility has been defined above.

**Discrimination index**

The discrimination index provides a measure of an item's capacity to differentiate more able students from less able students. On the basis of the total score on the test, and in relation to a body of knowledge or skills, students can be classified as being more able or less able. Typically, for an item with good discrimination, the more able students tend to choose the correct response whereas the less able students tend to choose one of the distractors. If a larger percentage of less able than more able students choose the correct (keyed) response, the item has a negative value for the discrimination index, and more probably than not, the question is unsuitable for general use.

A simple but quite satisfactory estimate of the value of this index for an item may be obtained in the following way:

> Arrange the test papers in rank order of total marks and divide this ranked set into approximate thirds. Disregard the middle group and, with the scores of equal numbers of students from the upper and lower groups, use the following formula for each item:

$$\text{Discrimination Index, DI} = \frac{R_u - R_l}{N}$$

where $R_u$ = number in upper group (1/3) who gave the correct response,

$R_l$ = number in lower group (1/3) who gave the correct response,

$N$ = number of students in either the upper or lower group.

NOTE: It is not necessary to choose exactly one third (1/3) the number closest to one third is quite satisfactory.

The possible range of the discrimination index is from -1 to +1. It provides an estimate of the correlation between the particular item and the results on the test as a whole. The more positive the value of the discrimination index, the more closely that item correlates with the items of the test as a whole.

*Consider the following two examples:*

*i) 32 students take a test and, for the purpose of estimating the discrimination index, equal groups of the top 10 scorers and bottom 10 scorers on the test are produced. For a particular item, 8 in the upper group and 3 in the lower group chose the correct response. The discrimination index for this item then becomes*

$$\frac{8 - 3}{10} = +0.5$$

*ii) On the same test, but for a different item, 2 of the upper group and 5 of the lower group chose the correct response. The discrimination index for this item is*

$$\frac{2 - 5}{10} = -0.3$$

It can be seen that the first item has a positive value (+0.5) and therefore is discriminating in the same direction as the whole test. However the second is discriminating negatively, that is, it is contributing negatively to the results of the test. Where this occurs, the test constructor should try to determine the reason. For example, the second item may be poorly worded or it may have two correct responses.

For small groups such as a class size of about 20, discrimination index values of greater than 0.4 are desirable, and anything less that 0.2 is unsatisfactory. As the size of the group increases, the value of the discrimination index which indicates a satisfactory item decreases, but for no sample size is a negative value or a value near zero acceptable. A good rule of thumb is to be suspicious between 0.3 and 0.2 and reject below 0.2.

More sophisticated techniques for the determination of discrimination indices are available but they take us beyond the needs of the classroom teacher. However, those interested in applying more accurate statistics should consult Exercise V(1). This exercise provides an alternative method for estimating the Discrimination Index for most classroom tests. It should be noted, however, that the above discussion on discrimination applies to situations in which a spread of scores is required, as in much achievement testing. This point will be taken up again in the section on **Statistics for criterion-referenced tests.**

**Distracter analysis**

In a multiple-choice item, the attractiveness of the distracters to the students is important. An item with a distracter which no one chooses or one with distracters which are too attractive, particularly to the more able students, should be examined for ambiguity. The keyed answer may not be the only correct response.

For most teaching purposes, a count of the number of times each response is chosen by the students provides a reasonable indication of how well the distracters are working. For a more complete analysis, particularly where the number of students taking the test is large enough, discrimination indices for each distracter can be calculated.

As a result of this type of analysis, items may be continually improved as more data become available. There is no point in retaining items for future tests if they invariably discriminate against the more able students; similarly, there is no point in retaining distracters that the students never choose.

**Storage of item statistics: norm-referenced tests**

Many teachers have found that their assessment procedures have been greatly improved by the systematic filing of item data. One simple way of doing this is by using a card system. A copy of the question is placed on one side of the card, and the relevant statistical information on the reverse side. An example of the data side of such a card is given below.

Subject: MATHS
Content: NUMBER – ADDITION
Cognitive objective: COMPREHENSION
Answer: B

| Test date | No. Tested | Total Number Correct | Upper/ Lower Groups | A | B* | C | D | Omit | Difficulty % | DI |
|---|---|---|---|---|---|---|---|---|---|---|
| 3/5/84 | 32 | 16 | Upper 10 | 0 | 9 | 1 | 0 | 0 | 50 | 0.6 |
| | | | Lower 10 | 1 | 3 | 2 | 4 | 0 | | |
| 4/4/85 | 35 | 14 | Upper 10 | 1 | 7 | 1 | 1 | 0 | 60 | 0.5 |
| | | | Lower 10 | 2 | 2 | 3 | 2 | 1 | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |

Such a filing system facilitates the ready location of items on required topics and provides a continual check on the efficiency of the items.

**Computer programs for item analysis**

Classroom teachers should also investigate the various shortcuts to the above procedures that are available in computer packages. Most computer companies offer a test analysis service that provides for on-going checks on individual items and a regular review of class performance.

Teachers are strongly advised to check what is available for the computers to which they have access.

Once a system has been established the time required to carry out item analysis is minimal and the potential rewards in terms of improved teaching efficiency are considerable.

## Statistics for criterion-referenced tests

The TAFE teacher is faced with a significant problem when it comes to statistics for criterion-referenced tests. Some of the techniques available require the teacher to possess a high level of mathematical sophistication. Other, simpler techniques, require the organisation of trial test procedures that are demanding of resources.

Two of the simpler procedures will be considered—pre-test:post-test and masters:non-masters.

Both procedures involve the validation of items by confirming that a period of instruction can result in students changing from 'novices' to 'masters'.

**Pre-test:post-test**

This method involves the teacher developing and administering a test prior to beginning instruction and then administering the same test at the end of the instruction period.

For any item the number of students getting the correct answer after instruction should be **significantly greater** than the number getting it correct before instruction. In the ideal situation 0% would get the correct answer before and 100% after. However, a much more likely result is that some would know or guess the answer beforehand and some would fail to get it correct after instruction. This failure could be through a lack of understanding on the part of the student, unsatisfactory teaching, or a combination of both.

It is therefore necessary for the teacher to decide what number of correct answers is significantly greater. This is essentially an arbitrary decision on the teacher's part. The critical number is, of course, the number getting the correct answer after instruction. In the criterion-referenced situation this must be high. If there are no unusual circumstances to take into account, anything less than 80% correct should be viewed with concern.

**Masters:non-masters**

This method is essentially the same as that of pre-test:post-test except two different groups of test takers are involved.

The teacher needs to be able to identify a group of individuals who are masters of the subject matter to be tested. Items are then tried out with a group of students before they have received instruction (non-masters) and the group of masters of the subject matter.

From the point of view of the item statistics this method has the advantage of being relatively independent of the quality of instruction. (For example, a 'good' item will not reveal itself as such in the pre-test:post-test situation if the quality of instruction has been poor). However, finding a suitable group of 'masters' presents formidable problems of organisation for most teachers.

**Statistical tests of significance for criterion-referenced tests**

As indicated earlier the mathematics of statistical tests for criterion-referenced tests is very challenging. Many of the approaches are still in the experimental stage. Those wishing to pursue this topic to a greater depth are referred to Livingston (1972), Berk (1980), and Swezy (1981).

**Storage of item statistics: criterion-referenced tests**

A card system similar to the one outlined on page 113 is a suitable way to maintain a record of the use of criterion-referenced test items.

It is also important to realise that once the validity of an item has been established it is not necessary to repeat the pre-test:post-test, or masters:non-masters validation exercise each time. A teacher can assume that the item will remain valid as long as there is no significant change in the content taught or the way it is being taught.

This is another advantage of storing item statistics as the teacher can note any marked departure in item difficulties from one test to another. When an item difficulty changes markedly between administrations the teacher should look for an explanation of the change.

## Validity via consensus moderation: a non-statistical alternative

Many TAFE teachers work in relative isolation and they can be placed in a position of having to make assessment decisions from a limited range of experience. However, as has been stressed in earlier chapters, the task of assessment can be made more straightforward if the teacher is working with a clear set of performance objectives stating:

- the skill to be performed
- the conditions under which it must be performed
- the acceptable standard of performance to be achieved.

It is also possible to further improve the assessment process if teachers can share their assessment decisions with colleagues. Discussions with other teachers about the work of students enables a consensus to be reached on what constitutes an 'acceptable standard of performance'.

In some States and Territories regular meetings of teachers are held to arrive at an agreement on skills, conditions and standards through a process known as consensus moderation. Teachers at these meetings discuss and agree on a common interpretation of the performance objectives. They bring along their students' work samples, the assessments of which are also the subject of discussion leading to consensus.

Consensus moderation can greatly enhance the validity of the assessment process. When the process does not occur as part of the formal system, teachers should try to set up their own informal network.

Other aspects of validity are taken up in the next chapter.

## REFERENCES


Berk, R. A. (Ed.). Criterion-referenced measurement: The state of the art. Baltimore, Md: John Hopkins University Press, 1980.

Izard, J. F. Construction and analysis of classroom tests. Hawthorn, Vic.: ACER 1977.

Livingston, S. A. Criterion-referenced applications of classical test theory. Journal of Educational Measurement, 1972, 9(1) 13-26.

Swezy, R. W. Individual performance assessment: An approach to criterion-referenced test development. Reston, Virginia: Reston Publishing Co., 1981.

Theobald, J. Classroom testing: principles and practice. Hawthorn, Vic.: Longman, 1974.

## EXERCISE V(1)

## THE PHI(Ø) CO-EFFICIENT AND ITEM ANALYSIS

**INTRODUCTION**

The Ø co-efficient is one measure often used to give the correlation between results on one item, and results on the test as a whole. On the basis of total test score, those tested may be allocated to an upper group or a lower group. It is convenient to let each group comprise one third, or one half, of the total.

The Ø co-efficient is then defined as follows:

| | item incorrect | item correct |
|---|---|---|
| upper group | B | A |
| lower group | D | C |

$$Ø = \frac{AD - BC}{\sqrt{(A+B)\ (C+D)\ (B+D)\ (A+C)}}$$

A chart can be used to make the calculation of Ø co-efficients for items on a test simple matter.

Consider the following example:

A class of 31 students was set the following question as part of a larger test.

Q. What is the maximum number of cartons measuring 200mm x 200mm x 100mm that will fit into a box 1m x 1m x 1m.

A 1 000 C 100

B 250 D 25

(Answer B)

The answers of the 31 students **on the whole test** were marked and totalled and the answer papers arranged in order from highest to lowest.

The papers of the top 10 and bottom 10 students were then compared for each question.

In the case of question Q the results were as follows.

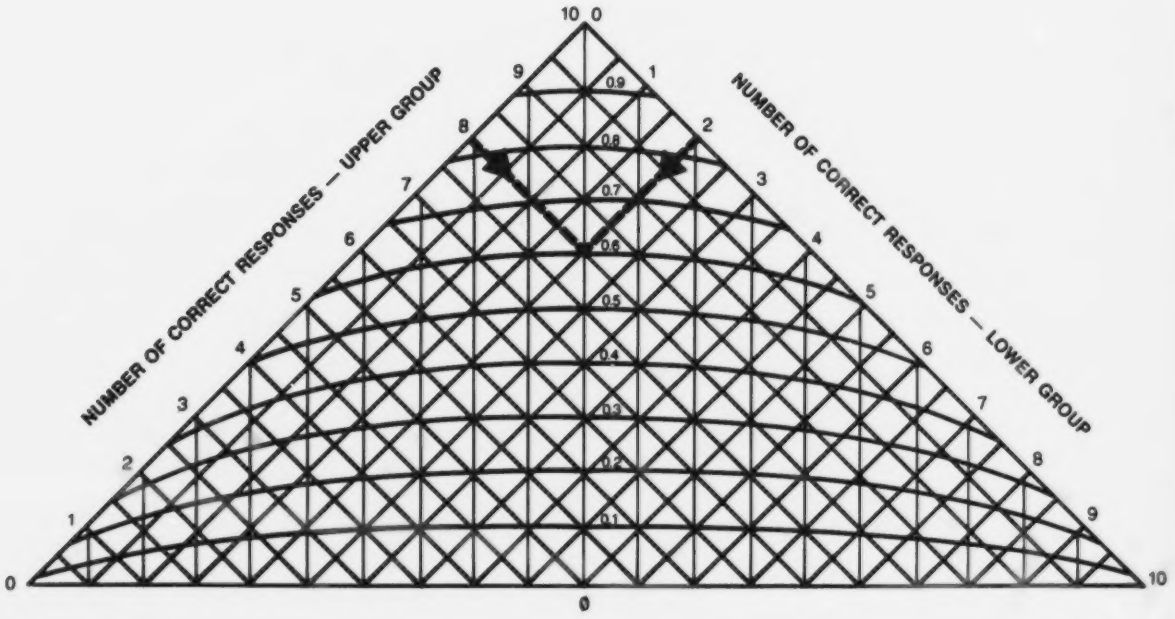
| | A | B* | C | D |
|---|---|---|---|---|
| Number from upper group (N=10) | 1 | 8 | 1 | 0 |
| Number from lower group (N=10) | 5 | 2 | 1 | 2 |

That is, 8 from the upper group answered correctly, but only 2 from the lower group.

The $\phi$ co-efficient can be obtained by applying the formula given above.

However a simpler method involves reading the value from the ABAC chart as shown below.

The lines from 8 in the upper group and 2 in the lower group intersect at $\phi$ = 0.6.




ABAC FOR THE DETERMINATION OF PHI(ϕ) COEFFICIENT

**EXERCISE**

Now consider the following item in a science test taken by a large group of 149 students.

Select the item out of place in the following:

A conduction of heat

B freezing of water

C filtration of sand from water

D burning of magnesium

E refraction of a beam of light.

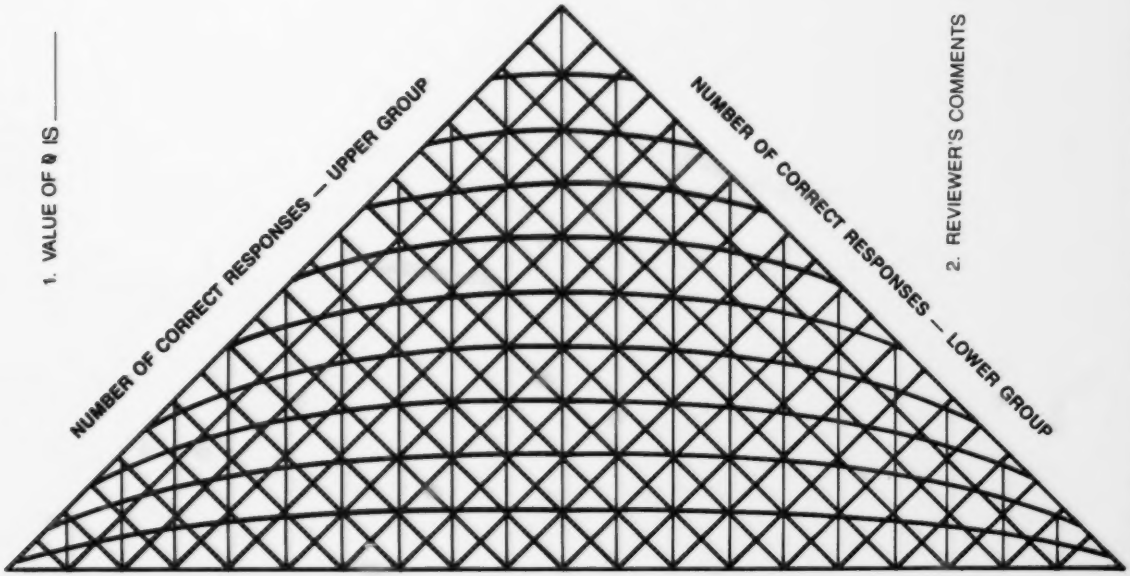
(Answer D)

After score the whole test the statistics for this particular item were:

| | A | B | C | D | E | N/A |
|---|---|---|---|---|---|---|
| Upper group N = 50 | 4 | 2 | 19 | 13 | 12 | 0 |
| Lower group N = 50 | 15 | 3 | 5 | 2 | 24 | 1 |

1. Calculate the $\Phi$ value using the ABAC chart provided below

2. Review the item in the light of the data and suggest ways to improve it, if appropriate.

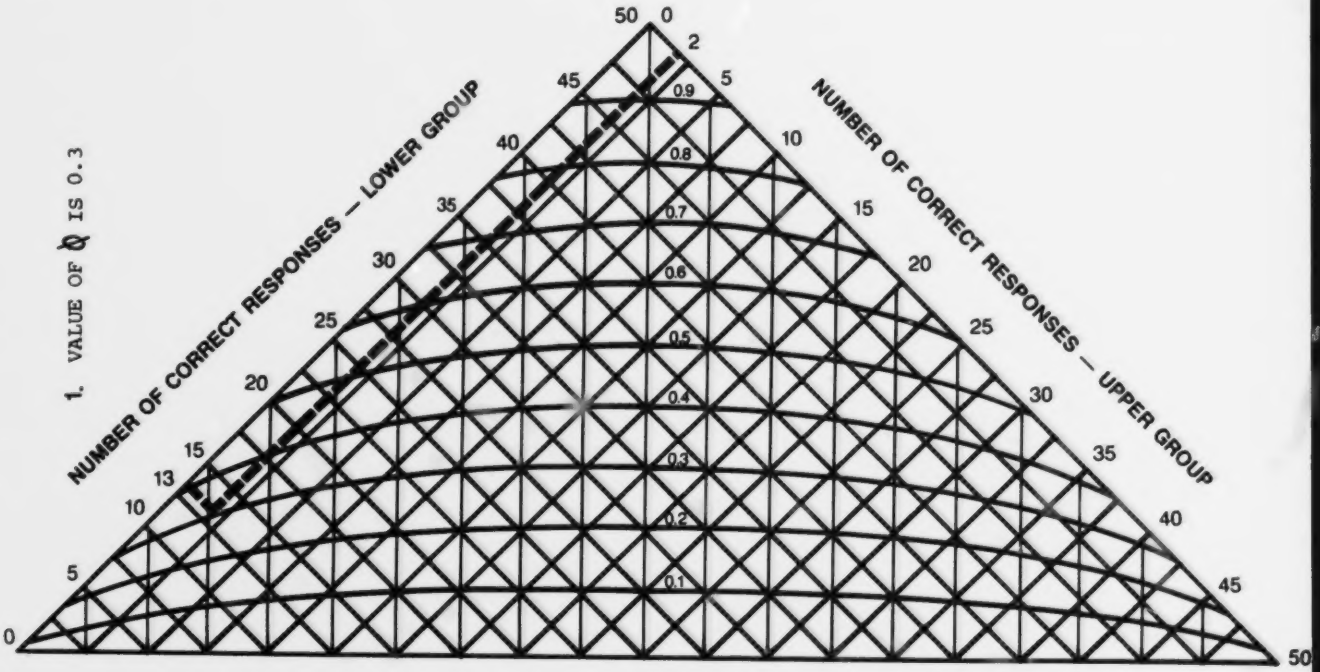



ABAC FOR THE DETERMINATION OF PHI(Φ) COEFFICENT

**SUGGESTED ANSWERS**

**EXERCISE V(1)**

1. VALUE OF $\phi$ IS 0.3




ABAC FOR THE DETERMINATION OF PHI (Φ) COEFFICIENT

1. VALUE OF $\phi$ IS 0.3

2. REVIEWER'S COMMENTS

NUMBER OF CORRECT RESPONSES—UPPER GROUP

NUMBER OF CORRECT RESPONSES—LOWER GROUP

ABAC FOR THE DETERMINATION OF PHI ($\phi$) COEFFICIENT

The question has problems.

Alternative C attracted a significant number of students from the upper group. It seems that they saw 'separation' of the two substances as sufficient reason to select C.

Since the question is about distinguishing between physical and chemical change the best solution would be to omit C altogether or to replace it with a less controversial example such as 'evaporation of alcohol'.

## CHAPTER VI

## TEST VALIDITY AND RELIABILITY

### WHAT THIS CHAPTER IS ABOUT

This chapter explains the importance of the properties of test validity and reliability as they apply to classroom tests.

*OBJECTIVES*

*After completing this chapter you should be able to:*

- *explain the meaning of the terms validity and reliability and distinguish between them.*
- *describe three different types of validity of importance in classroom testing.*
- *explain the sources of error in tests and how these relate to reliability.*

### INTRODUCTION

Chapter II explained the importance of objectives in test construction and how test items are written to assess (measure) objectives. When a teacher writes items which appropriately assess an objective the items are said to be **valid** measures of the objective.

Should the teacher then use these items with his or her classes on a number of occasions over a period of time it is possible to obtain an estimate of their reliability. To be classed as **reliable**, repeated testing with similar groups of students should produce approximately the same item difficulties.

The properties of an item or test that make them valid and reliable will be explored in more detail in the following sections. However, in summary, **validity** is concerned with how well a test meets its stated purposes (objectives); **reliability** is concerned with the accuracy of the test as a measurement instrument.

The TAFE teacher must be concerned with both these qualities of a test but considerations of validity are particularly important in a classroom situation. It is most important that the teacher be confident about the validity of a test, however the situation with reliability is different. It is possible to have a highly reliable test which is invalid for classroom use.

*For example, a mathematics test consisting of items which require the addition of two numbers, each consisting of two digits, such as*

| | | | | |
|---|---|---|---|---|
| *23* | *51* | *73* | | *91* |
| + | + | + | | + |
| *42* | *29* | *86* | ..... | *19* |

*would likely prove to be highly reliable when used with students taking a subject such as Trade Mathematics.*

*However, if the students taking the test were not taught addition of two digit numbers or never used this computational skill in their trade, then the test has little or no validity for them.*

Symbolically the relationship between validity and reliability can be illustrated with the results of a shooter firing at a target.

*Each time the bull's eye was the target of the shooter, yet the final result was a grouping of holes off to the side as shown in the diagram.*

RELIABLE BUT NOT VALID

*The shooting was consistent but the shooter has not got the result wanted.*

Teachers must avoid such mistakes, validity is the prime concern in classroom testing.

**VALIDITY**

To say a particular test is 'valid' can be misleading. Tests cannot be termed valid by themselves—they have to be valid for something such as, 'a valid test of reading comprehension' or 'a valid test of knowledge about legislation in the building industry'.

*Furthermore, a single test can be valid in a number of ways. Consider, for example, a maths test used as a screening test of applicants for places in a course. The same test could also be used to decide whether participants already in a course need some remedial assistance in mathematics. The validity of the test is quite different in the two cases.*

*Test validation concerns the interpretations made of the results of the test.*

In this section we will consider three types of validity:

- face validity
- content validity
- predictive validity.

**Face validity**

It is possible to read the items in a test on plumbing and make a judgment as to whether or not they are relevant to a plumbing course. If it seems as if they are testing relevant knowledge, skills and attitudes then it can be said that the test has **face validity.**

The important point to appreciate about face validity is that the judgments involved are superficial and may not stand greater scrutiny. Some test constructors are critical of the notion of face validity and suggest it could be more appropriately termed 'faith validity'.

Nevertheless all tests should appear to be testing what they are, in fact, testing because it gives the person taking the test greater confidence. However, judgments about validity should be more than superficial.

**Content validity**

When the items in the test are written to assess the performance of students on the objectives of a subject, such a test has **content validity.**

Content validity is of prime importance in most TAFE courses. For the purposes of this section it will be assumed that the curriculum presented to the students has 'content validity'.

*For example, we will assume the curriculum for a course such as the TV Technicians Certificate presents everything that students need when they work as TV technicians after having successfully completed the course.*

Our concern is with the validity of the tests used in the courses, the validity of the curriculum will not be considered.

The task of the teacher is to select content and to devise test situations which accurately reflect the objectives of the course. In devising the test situation the teacher must also clearly specify the rules that are to be applied for judging student performance.

> *For example, a test for TV technicians would obviously include the repair of parts of a television set. However to be consider as having 'content validity' such a test would need to be accompanied by a clear set of instructions and state a time limit in which the task had to be completed.*
>
> *Successful completion of a TV Technicians course involves being able to carry out repairs within certain time limits—and this is part of the content validity of the test.*

Content validity is largely about judgments. The judgments are mainly concerned with what will be externally observable as a result of of testing. Furthermore, the judgments are made without considering the persons being tested. It is the task that is all important. If allowances are made for individuals, content validity is at risk.

> *In the case of our TV technician students used above, it is perfectly valid to fail a student who takes 30 minutes to repair a part that the curriculum specifies should be repaired in 5 minutes.*

The test grids referred to in Chapter IV are examples of the method by which a teacher can ensure that a test has high content validity. In such cases the test must adequately sample the subject matter and performance objectives and reflect the ways in which the various elements have been taught.

These are all judgments which the teacher is able to make. Indeed, the teacher is the person best placed to make judgments about content validity because not only does he or she know what content has been taught, but also how much emphasis has been given to particular elements.

In Chapter V, the validity related to the procedure of consensus moderation was content validity.

## Predictive validity

On many occasions we are interested in now students will perform at some time in the future. Many tests are constructed with future performance very much in mind. A good example is the Year 12 examination which is used to select students for places in tertiary institutions. Here selection is based on the assumption that a student who does well in the Year 12 examinations is more likely to be successful at tertiary studies than a student who does not do well.

The Year 12 examination is said to have **predictive validity** because it predicts success in the future. (In actual fact the predictive validity of the Year 12 examination is not good, however a certain 'folklore' has become associated with the examination and people perceive it as a good predicter despite the evidence to the contrary.)

Any claims about predictive validity are, of course, testable and can be measured by seeing how well the performance on the test correlates with performance in the future.

TAFE teachers are not generally required to give consideration to predictive validity when framing their tests. One point to bear in mind, however, is that successful completion of a subject or course usually implies something about future performance.

When a teacher says that a student has 'passed' the course in carpentry and joinery this means the teacher is making a judgment about how that student will perform in the trade. The teacher's tests which enabled the 'pass' decision to be made therefore have a predictive element in them. From time-to-time teachers should follow up some of their ex-students and try to reassure themselves that their judgments are being confirmed in the workplace.

### Other types of validity

There are a number of other types of validity but they are unlikely to be encountered by many TAFE teachers. Those wishing to pursue the subject should consult the standard texts, e.g. Thorndike (1971).

## RELIABILITY

The reliability of a test is concerned with the test's accuracy as a measurement instrument.

All measurement involves error and it is the size of the error that determines the reliability of the measure.

In education, the scores obtained in tests can be in error for many reasons. We know, for example, that assignments are likely to receive different scores from different markers. Furthermore, the same marker when asked to mark an assignment for a second time may well give it a different score on the second marking.

Error can also occur when students are asked to repeat the same test. Students may perform differently on the second occasion because

- they have learnt (or forgotten) something in the period between the two testings
- the test conditions have varied (e.g. noise level, weather, lighting)
- their state of health has changed (e.g. migraine attack, influenza)
- their motivation has changed.

Teachers can also make mistakes in adding and recording marks.

The more error that occurs in a test the more unreliable that test is.

The reliability of a test can be measured in a number of ways. For example, in cases where it is possible to test a group of individuals on two occasions, a reliable norm-referenced test would generally rank the individuals in almost the same order on each occasion. By calculating the correlation co-efficient between the two groups a **reliability co-efficient** can be obtained. Such calculations are not applicable in classroom teaching situations, however teachers should be aware that longer tests with more items have a greater reliability than shorter tests.

In the case of criterion-referenced testing, where ranking is not appropriate, most individuals should score the same mark or grade on the two occasions they take the test if it is to be considered reliable.

## REFERENCES

Thorndike, R. L. (Ed.) Educational measurement (2nd ed). Washington: American Council on Education. 1971.

## CHAPTER VII

## REPORTING RESULTS

### WHAT THIS CHAPTER IS ABOUT

The reporting of results in TAFE is presented as an area which deserves much more attention than it presently receives.

Contrasts are drawn between what are the legitimate needs of the clients of the TAFE system and what they at present receive by way of reports.

Suggestions are made as to how the present system of reporting can be improved.

*OBJECTIVES*

*After completing this chapter you should be able to:*

- *identify the different audiences for TAFE reports*
- *explain the needs of the different audiences for TAFE reports*
- *identify the necessary components of a good reporting system*
- *explain the use of profile reporting.*

### INTRODUCTION

A chapter on reporting results in TAFE would seem, at first glance, to have a relatively small amount of information with which to deal.

TAFE colleges in common with Universities and Colleges of Advanced Education, generally provide only limited information on student achievement levels after the completion of subjects and courses. Furthermore, they provide almost no information about the progress of a student who is in the process of taking a particular subject. This level of reporting contrasts markedly with that available in the secondary school system where it is usually possible to get term by term (or half-yearly) reports of a student's progress on a subject-by-subject basis.

This chapter will describe a number of improvements that TAFE classroom teachers might seek to implement in their reporting methods. These improvements are not designed to bring TAFE more in line with secondary schools, but rather to look at new ways of reporting that are appropriate to technical and further education.

As a first step it is necessary to consider the audience the reports are meant to serve and what information a report should provide.

## AUDIENCE FOR REPORTS

Different individuals and groups have a legitimate interest in a student's report. The three most important are:

- the student who is the subject of the report
- the parents of the student
- employers.

Some might argue that the student is the only one to whom a report should be directed. However, this fails to take account of the stake that others might have in the results of the student's education and training. For example, it is likely that the parents who have provided financial support will believe they have a right to check on their investment, or employers who have jobs that need filling will want help in making their selections. Employers can rightly argue that it is also in the students' interest for the most suitable applicants to be chosen for jobs. Therefore, while it is hard to deny that the student must be the first concern of the report writer, the concerns of other potential audiences should not be overlooked.

### The student

A number of procedures have evolved in recent years aimed at protecting students from being unfairly disadvantaged by their reports. One way of doing this is to give students the control over the release of the information. This means that reports cannot be released to other parties, for example, a potential employer, without the written permission of the student. Just how effective such a procedure is in practice is open to question. Employers are likely to think twice about offering a position to an applicant who refuses to let them have access to his or her report.

Another method is to allow students to have an input into their own reports. This can be done either by the student and teacher jointly deciding what should be reported, or by getting the student to comment in writing upon the teacher's report.

#### Value of reports to student

A number of studies have shown that evaluative feedback is positively related to increased academic performance by both high and low ability students (Kash and Borich, 1982). Feedback can vary in its value—that which occurs during a course of study is clearly able to affect the students' on-going performance. For example, a report which states there is a need for improvement should motivate a student to produce better work. Feedback at the end of a subject has less value but it can help students plan for the next segment of their course.

The value of reports to students therefore centres on feedback and its consequent motivating influence. However, teachers should be aware that motivation is a delicate issue and where reports contain insensitive criticisms the result may well be to depress student motivation.

It has already been pointed out that despite the fact that feedback in the form of a written report is of value to students there is little of it occurring in TAFE. This point is taken up later.

## The parents

The parents of TAFE students are seen in a different light from parents of secondary school students when it comes to reports. In some schools the reports may actually be written to the parents through comments such as:

> *You can be assured that the school is doing all it can to improve John's performance in this subject.*

TAFE students, however, are treated as mature individuals who are more in charge of their lives, and parents are therefore a secondary audience.

### Value of reports to parents

As indicated earlier, parents can have a legitimate interest in students' reports if they are providing financial support.

At the basic level parents want to be informed about the success or failure of their offspring, but their needs are, in reality, usually greater than this.

Parents are usually only vaguely aware of the objectives of the subjects and course their children are taking. Therefore, if a report is to adequately inform parents, it needs to report in terms of the objectives the teachers set out to achieve. It also needs to be free of educational jargon so that it can be readily understood.

It is therefore necessary to provide a statement of the aims of the course. This should not be a detailed syllabus document but rather a summary outline that is easy to read. Such statements can be printed on the back of the report form.

For example, consider the following statement for a Clerical Pre-vocational Course.

*CLERICAL PRE-VOCATIONAL COURSE*

*Communication, operative and manipulative skills*

*The student can demonstrate competence in:*

*Handling mail (incoming and outgoing)*

*Parcelling, addressing, weighing, stamping, sorting and distributing mail.*

- *Use of a post-book*
- *Use of a franking machine.*

*Duplicating and copying*

*Use of carbons, ink duplicators and photocopiers, and hand collation.*

- *Use of a spirit duplicator*
- *Use of a collating machine*
- *Use of electrical and heavy-duty staplers.*

*Filing*

*Using the filing systems which are commonly available.*

*Record keeping*

*The maintenance of records using dockets, forms, invoices, books and stock record cards.*

*Using records as a source of information.*

*Machine operations*

*Use of a typewriter for everyday operations, such as typing envelopes, simple memos and forms.*

- *Use of simple accounting machines, such as ad-listers, tills.*

*(Adapted from Royal Society of Arts [UK] Vocational Preparation Clerical Course)*

## The employers

There is a relatively high degree of compatability between the interests of employers and TAFE students over the matter of reporting results. Nevertheless employers frequently criticise the reports they receive for lacking important information. However a problem this creates for the TAFE teacher is that it has been shown the information employers consider to be important varies from job to job. The students' educational achievement is almost always seen as necessary information but Wilson (1982) has reported a considerable variation after that. Items identified in the Wilson survey of Victorian employers that are of particular interest to TAFE teachers include:

- the ability to communicate
- interest in the job
- the ability to work well with others
- willingness to be flexible in job
- the desire to do well
- possessing initiative
- reliability.

### Value of reports to employers

Employers value reports which provide valid assessments of their employees or potential employees. Good reports save employers time and money that they might otherwise need to spend on developing and operating their own assessment procedures. As with the case of parents mentioned earlier, for a report to be of value, employers also need to be aware of the teaching objectives of the subjects reported and to have reports free of educational jargon.

## Other audiences

Two other groups should be mentioned as legitimate audiences for reports namely:

- teachers and
- admissions officers of other tertiary institutions.

Teachers have a dual interest. On the one hand the reports of student assessments provide an indication of the effectiveness of their own teaching and are therefore likely to have a motivating effect on them in much the same way as they do for their students.

Teachers also value reports that have come from other teachers as they help them to understand the strengths and weaknesses of their charges. Reports are particulary valuable in the case of students transferring from other colleges.

In a similar way admissions officers of other tertiary institutions often require a range of information to assist them in deciding about selection and placement of students who are applying for entry with either a full or part qualification from a TAFE college.

## REPORTING MARKS AND GRADES

Teachers usually give marks to pieces of work they are assessing. In the criterion-referenced approach common in TAFE subjects it is a relativly straightforward matter to give a mark of 1 when an item on a specific objective is done or answered correctly (mastered) and 0 when it is not.

All that then needs to be decided is how many such items a student needs to do or answer correctly in order for the teacher to be confident that the specific objective has been mastered. This problem was discussed in Chapter V where it was suggested that 80% mastery was appropriate in most cases—that is, 4 out of every 5 items done correctly.

On the other hand, should the teacher be using a norm-referenced approach to assessment, marks might mean something different. A teacher who sets very high standards might give a mark of 50% to a student who he or she believes has mastered a task (e.g. writing an essay) and progressively higher marks for better and better work. The problem here is that unless we know the teacher's standards we do not know what the marks mean.

This difficulty is related to the problem of interpreting reports referred to earlier. Not only is it necessary to tell the reader of a report what each teachers objectives are, but it is also necessary to explain the meaning of any mark or grade that is given.

In some subjects letter grades (e.g. A to E), are commonly given instead of marks. These enable some of the worst abuses of marks to be overcome. If a group of students are awarded an A it means that their work is of very high standard and meaningless within-group comparisons are avoided (e.g. suggesting that a person with 99% is superior to another with 97%).

Unfortunately it does not avoid almost equally meaningless between-group comparisons. However, perhaps the greatest value of reporting letter grades is that they can inhibit the reporting of composite scores.

*The practice of adding the marks of different Year 12 subjects together and claiming to have something which is meaningful is a widespread Australian practice. Yet a person who buys 50 strawberries and 10 apples does not think of the purchase as being 60 pieces of fruit or 60 'strapples'. But, if John gets 50 out of 50 for a Mathematics test and 10 out of 50 for English some people have no trouble in seeing this as 60% overall for Maths and English!*

*In fact, John would seem to have a flair for Maths while experiencing difficulties with his English, and the 60% completely masks these important pieces of information.*

The tertiary entrance authorities around the country seem to have given little thought to the consequences of their continued use of composite scores of the Year 12 examinations.

Of course, using a criterion-referenced approach to assessment, and reporting in terms of performance on objectives, will largely overcome the problem outlined above.

However even when results are reported for a single subject there can be difficulties. For example many TAFE subjects are mixtures of theory and practical. Here the temptation might be to allow high performance on theory to compensate for low performance on practical or vise versa.

Such an approach has implications similar to those already discussed and teachers should avoid falling into such traps. The task of teachers is to ensure that satisfactory completion of a subject really means just that and once they have set their standards (be they norm- or criterion-referenced based) these are the standards to which they adhere.

**PROFILE REPORTING**

The idea of student profiles has been around for a long time. Put simply, a student profile is a document produced by teachers and trainers which provides information about a student's knowledge, skills and attitudes relevant to a particular curriculum.

Profiles have developed from attempts to provide more information about a student than has been available through the traditional examination system.

The need for more information about students has been a continuing plea of employers. Indeed, all involved in the transition from TAFE to work can benefit from more information, and profiles are one way of providing that information.

Although profiles are being used increasingly it is important to note that their effectiveness remains poorly researched. Two United Kingdom publications (Further Education Unit 1982 & 1984) cover a wide range of issues related to the state of the art of profiles development and give details of their strengths and weaknesses. Among the research evidence still being sought is an answer to the question 'How valid are profiles'?

While they are basically very supportive of profiles, the Further Education Unit publications point out that the subjective nature of the reporting of the attitudes and other personal qualities of individuals remains a problem. There are also questions about the level of acceptance that profile reports has among the wider community.

Two examples of profiles developed in the United Kingdom follow. Australian TAFE teachers are encouraged to try out this approach in their own classrooms and workshops.

## A profile for adult instructors and supervisors

The following profile was developed by the City and Guilds of London Institute and the Manpower Services Commission.

The profile describes the course performance of men and women working with unemployed 16-19 year olds who are taking part in the UK's Youth Opportunities Programme (YOP). Five basic areas of skill are provided by the course, namely:

Instructional Skills
Communication Skills
Caring/Guidance Skills
Evaluation Skills
Job Skills.

The content of the course is based on a set of 22 prescribed objectives covering each of the five areas of skill.

*A sample Profile Report is provided on pages 142 to 144. Each objective has its own set of descriptors showing possible levels of achievement of the objectives. Minimum acceptable levels of achievement are shown by the HIGHEST level of hatching on a descriptor. For example in objective 6, (page 142) and level 1 the descriptor: 'Pays only some attention to the importance of these factors' is the minimum acceptance level, whereas in objective 9, (page 143) the level 2 descriptor: 'Makes a fair effort to encourage communication' is the minimum acceptable level.*

The assessment of the achievement of the objectives is carried out over the full period of the course. Some of the objectives can be assessed during the taught part of the course. For example, objective 6 can be assessed through the course participant performing in a class exercise, case studies or role-playing. Objective 11 can be assessed though observation of discussion or a short piece of written work. Other objectives, for example 20 and 22, are best assessed by observing the Instructors/Supervisors at their place of work when they are working with trainees. In order to make an accurate assessment of objectives 18 and 19 the course tutor will consult with the YOP scheme manager to whom the Instructor/Supervisor is responsible.

The objectives and their descriptors are intended to be used by the course tutor and the course member together to enhance the dialogue, analysis and assessment relevant to the learning experiences of the course member.

On the last day of the course each course member is interviewed and their final (summative) profile is agreed between the course member and the course tutor(s). The tutor's observation in the form of a supportive written comment is usually agreed at the same time.

On the example shown on pages 142 to 144, the final agreed assessment levels are shown by a double line being drawn vertically in the appropriate descriptor box.

The course tutor indicates how well the course member has performed within the individual descriptor by positioning the double lines to the left or right of the centre of the descriptor box as is appropriate.

Further details of this profile approach can be obtained from Profiles (Further Education Unit, Sept. 1982, pp. 12-18).

*Fig. 1*

**PROFILE – INSTRUCTIONAL SKILLS** NAME: Phil Woodward

| Objectives The trainee instructor/supervisor should be able to: | Profile Grade 1 | 2 | 3 | Tutor's Observations |
|---|---|---|---|---|
| 1 Identify and select appropriate instruction methods and plans for instructing individuals and/or small groups. | Identification and selection is restricted in most cases – limited grasp of methods available. | Identification and selection is correct in most cases – some lack of variety. | Identification and selection is correct in all cases. | |
| Mark present status | | | | |
| 2 Make effective use of available resources for updating his own knowledge, and preparing instructing/-learning materials. | Makes only partial use of available resources, and even then in a limited way. | Makes use of easily available resources, without seeking further afield. | Makes full and effective use of all resources for planning and preparation. | Off the job training sessions shared best use of resources |
| Mark present status | | | | |
| 3 Select and use appropriate resources, including audio-visual materials, effectively in a range of instructing/-learning situations. | Uses learning aids and resources in a very limited way. | Makes use of available resources, but uses limited creativity. | Uses resources and aid to the fullest effect in the learning situation. | |
| Mark present status | | | | |

**PROFILE – COMMUNICATION SKILLS** NAME: Phil Woodward

| Objectives The trainee instructor/supervisor should be able to: | Profile Grade 1 | 2 | 3 | Tutor's Observations |
|---|---|---|---|---|
| 4 Recognize problems experienced by his trainees in written and oral communications, and offer realistic help. | Recognizes some problems, but has difficulties in offering remedies. | Recognizes several, but not all problems; will succeed at most attempts at correction. | Makes full identification of all problems, and offers realistic and helpful solutions. | |
| Mark present status | | | | |
| 5 Demonstrate different ways in which effective communication may be achieved. | Has some difficulty in appreciating different methods of effective communication. | Uses a fair range of methods, occasionally inappropriate. | Good demonstration of a wide range of communication methods. | |
| Mark present status | | | | |
| 6 Identify the purpose and importance of vocabulary, accent, style and form in communication. | Pays only some attention to the importance of these factors. | Recognizes the importance of some factors, but does not use them fully. | Stresses fully the importance of these factors, and uses variations as appropriate. | |
| Mark present status | | | | |
| 7 Identify appropriate forms of speech and writing for use in his own instructing/ supervisory situations. | Has some difficulty in choosing appropriate forms for day-to-day occasions. | Chooses appropriate form on most occasions but not all – some difficulties in style. | Uses realistic and appropriate styles on all occasions. | |
| Mark present status | | | | |

**PROFILE – COMMUNICATION SKILLS – (Cont'd)** NAME: Phil Woodward

| Objectives The trainee instructor/supervisor should be able to: | Profile Grade 1 | 2 | 3 | Tutor's Observations |
|---|---|---|---|---|
| 8 Use a wide range of verbal, non-verbal and visual methods of communication in his own instructing/ supervising situations. | Relies on a limited range of methods, with little understanding of student difficulties. | Uses a moderate range of methods, but lacks some skill in modifying them to meet student need. | Uses a full range of communication methods effectively to the students' advantage. | |
| Mark present status | | | | |
| 9 Demonstrate ways of helping his trainees to communicate effectively. | Makes only a limited effort to help trainees with communication problems. | Makes a fair effort to encourage communication. | Makes a sustained effort to encourage trainees to communicate effectively. | Communicates well with trainees – shows the importance of different types of communication |
| Mark present status | | | | |

**PROFILE – CARING/GUIDANCE SKILLS** NAME: Phil Woodward

| Objectives The trainee instructor/supervisor should be able to: | Profile Grade 1 | 2 | 3 | Tutor's Observations |
|---|---|---|---|---|
| 10 Describe fully the role of his own institution in the provision of training opportunities for young people. | Description is limited, with inaccuracies, and expressing an incomplete view. | Description is generally accurate but lacks detailed information. | Gives full and accurate description, relevant to the institution in question. | |
| Mark present status | | | | |
| 11 Describe the nature and problems of employment and unemployment, relative to his trainees. | Presents a shallow amount of information without much helpful detail. | Description is rather generalized, but includes some relevant and important details. | Presents clear, accurate and helpful information, relevant to the trainees. | |
| Mark present status | | | | |
| 12 Offer initial help and guidance to his trainees in relation to career, education and personal problems. | Only offers occasional guidance, somewhat removed from trainees' problems. | Plays a limited guiding/caring role. Gives helpful facts but has some problems at a personal level. | Plays a comprehensive role as advisor. Shows a full appreciation of trainees' problems. Helpful in all respects. | Prepared to accept the role of a first line adviser |
| Mark present status | | | | |
| 13 Advise his trainees about alternative sources of help and guidance, if needed (referral system). | Refers most cases to other people right from the start, or endeavours to advise comprehensively even when not qualified to do so. | Makes some use of referral system – has some difficulty in selecting appropriate additional help. | Uses referral system well. Shows interest by further enquiry. | Uses referral system to advantage |
| Mark present status | | | | |
| 14 Encourage and assist trainees to adopt a positive attitude towards their work and their adult roles. | Not very encouraging. Gives a little assistance if the need is pressing. | Gives encouragement and assistance on occasions, but is influenced by individual trainees and their needs. | Gives full encouragement and assistance to trainees. | A difficult area at first for Phil, but good team spirit is being developed |
| Mark present status | | | | |

**PROFILE – EVALUATION SKILLS** **NAME:** Phil Woodward

| Objectives<br>The trainee instructor/supervisor should be able to: | Profile Grade<br>1 | <br>2 | <br>3 | Tutor's Observations |
|---|---|---|---|---|
| 15 Take part in setting performance goals for the trainees, based on the objectives of the organization and the trainees' skills.<br>Mark present status | Takes only a limited part due to lack of required skill or involvement. | Modest participation using only a limited range of goals. | Fully participates in a realistic goal-setting exercise. | Goals based mainly on practical skills and job knowledge |
| 16 Choose or devise suitable assessment and evaluation methods for (a) checking trainees' attainment and progress, (b) discovering their learning problems, and (c) checking own effectiveness as a supervisor/instructor.<br>Mark present status | Makes limited effort to assess and evaluate using simple methods. Needs additional skills. | Falls short of a full appreciation of the techniques for assessment. Tendency to rely on subjective methods. | Chooses and/or devises appropriate methods at all times for the three areas shown. | |
| 17 Control the trainees' learning situation through effective feedback and appraisal, and by meeting changing needs. | Loses some control by using methods which are inappropriate. | Uses appropriate methods on some occasions. Limited control of the learning situation. | Uses appropriate methods at all times. In full control of the learning situation. | Best done in all cases through practical exercises |
| **18 Construct and present oral and written reports on trainees.**<br>**Mark present status** | **Produces reports which are inconclusive and badly organized.** | **Constructs and presents reports which are incomplete, lack final polish and have only restricted value.** | **Constructs and presents clearly understandable and worthwhile reports.** | |
| **19 Operate satisfactorily in a clerical/ administrative system.**<br>**Mark present status** | **Lack of liaison with the system – grudging use of support services.** | **System used haphazardly – sometimes fails to supply with needed information.** | **Support systems used correctly at all times.** | |
| 20 Demonstrate appropriate skill, knowledge, competence and awareness of safe practice in an area of work.<br>Mark present status | Does not demonstrate appropriate knowledge and competence in an area of work. | Has limited knowledge and competence. Needs some updating. | Demonstrates appropriate knowledge and competence in an area of work. | A first class craftsman, with sound knowledge. Sets a good example. |
| 21 State the extent of his own responsibilities in ensuring satisfactory and safe conditions of work for his trainees.<br>Mark present status | Statement is incomplete because knowledge is poor with inaccuracies. | Statement is generally acceptable, but lacks detailed knowledge. | Makes full and accurate statement relating to his institution. | Tries to make trainees aware of the importance of safe working conditions. |
| 22 Plan and organize within his area of work, to enable trainees to be most effectively occupied.<br>Mark present status | Relies on very limited planning and organizing of work. | Plans and organizes work but sometimes fails to use them. | Plans and organizes work to fullest effect. | Given the constraints of his working situation, he plans his work well. |

## A profile for a pre-vocational course

The second example of a profile, one used with pre-vocational training schemes, was also developed by the City and Guilds of London Institute.

This profile has four components as shown in the diagram.

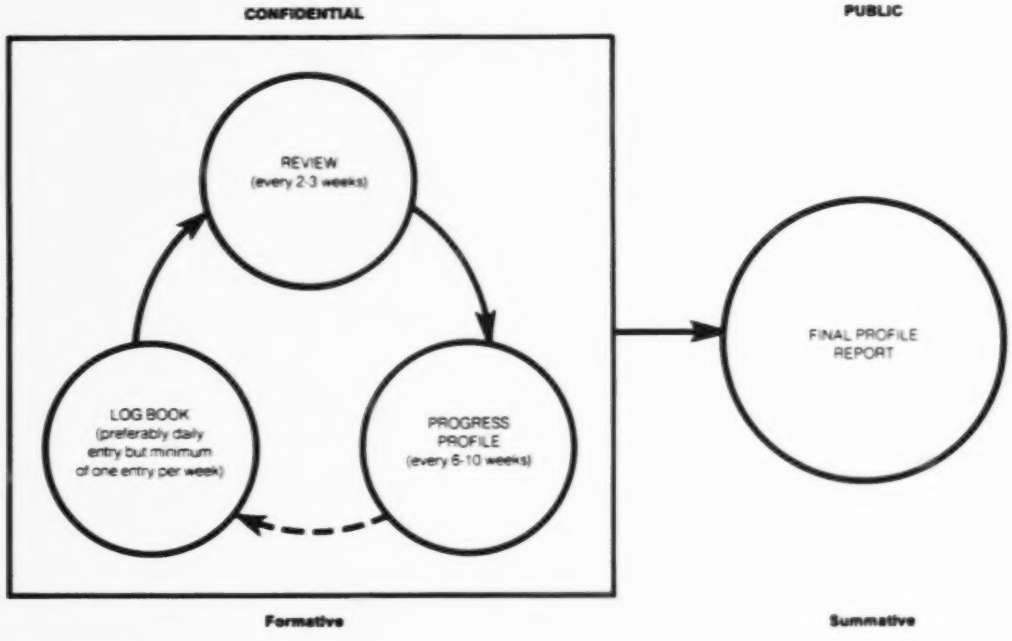



**Figure 2** The CGLI Profiling System

The first three components make up the formative system, the purpose of which is to support the student's learning by ensuring that their needs are identified and that the learning program is satisfying those needs. These components are confidential to the student and to the tutor. The fourth component, the final (summative) profile report, serves the entirely different purpose of communicating the student's attainments to a third party, for example, an employer. It is, therefore, a public document.

### Log book

The Log Book which is kept by the student, provides a record of what she/he has learnt and experienced and forms the main input to reviews.

An example of the two pages of a Log Book entry for a particular week follow - the entry relates to a student undertaking a pre-vocational nursing program.

Figure 3

**WEEKLY RECORD OF TRAINING**

WEEK COMMENCING .....

WEEK NO. 4

1. What work did you do and where?

Monday Bathed had a meeting with Staff and discussed with Mr Leahey my project Tuesday Bathed fixed box for Christmas hamper.
Wednesday went to College
Thursday Shaved John G and Bob F in morning, took some residents out for a walk around the building. John F came home from hospital
Friday Took Jessie to dentist. Have to make Sure John F gets enough to drink.

2. What did you do that was new?

The Staff meeting was new, Shaving the men was new, taking Jessie to dentist and having to watch that John takes enough liquid because of his kydneys.

3. What Off the Job Learning/Further Education did you undertake?

Learnt about Health and Safety Act and about Giro accounts.
Practised some first aid.

**Figure 3** continued

4. Was there anything you especially enjoyed during the week and why?

Taking some of the residents out for a walk was great because I had two residents with me and one complained of the cold and the other tried to tell her that she was talking nonsense, and it was good to see who would give in just. None did. I just took them in.

5. Was there anything you did *not* enjoy and, if not, why not?

Going to the dentist as we were just in about two minutes after waiting quarter of an hour and I didn't like the dentists attitude to the resident involved.

6. Supervisors comments.

This week Gillian has progressed well.
Used her initiative with residents and certainly engaged herself. Has also been helping with Xmas preparations.
This has been quite a varied week for Gillian but she has coped very well.

Trainee's Signature Gillian Anderson Date 14-10-83

Supervisor's Signature J Reilly Date 14-10-83

CGL/00435. S25. © 1983. City and Guilds of London Institute

## Reviews

Reviews enable student and tutor to discuss progress and to make plans for future work. They are also intended to be useful for identifying and resolving problems and for building good working relationships. The review, therefore, is at the heart of the formative profiling process. A jointly agreed record of each reviewing session is maintained within the log book.

An example of a review follows.

Figure 4

**FOUR WEEKLY REVIEW OF PROGRESS**

1 What are the main tasks or work areas covered in the last four weeks?

Care Work - bathing

Kitchen and domestic

Laundry

2 What has the trainee done particularly well?

All aspects of caring. Has a pleasant reassuring nature.

Has worked well with clients on hospital visits.

3 After discussion, we have agreed progress has been made in:

Integrating with the staff. Has developed in self confidence.

but attention needs to be paid to

Projecting herself. Becoming a little more self assured.

Contributing at staff meetings.

4 The next month will be spent

Working on project "Confused Elderly."

More involvement with baking etc. in Kitchen.

Trainee's Signature Gillian Anderson Date 14-10-83

Supervisor/Tutor Signature J Williams Date 14-10-83

CGL/00436. S25. © 1983 City and Guilds of London Institute

**Progress profile**

A Progress Profile is completed jointly by the tutor and the student every six to ten weeks. It records specific applications of abilities, the examples being drawn from the student's work and experience, thereby setting a context for the 'Progress in Abilities' statements on the grid. The 'examples' stage is always completed first as it provides the basis on which to judge the level of ability which has been achieved. This level is recorded by shading-in the bar up to and including the appropriate level on the grid.

Page 150 shows an example of a completed Progress Profile.

BEST COPY AVAILABLE

**City & Guilds of London Institute**

**Figure 1**
**Progress Profile**

| | ABILITIES | EXAMPLES OF ABILITIES | PROGRESS IN ABILITIES | | | | |
|---|---|---|---|---|---|---|---|
| COMMUNICATION | TALKING AND LISTENING | Listens to patients sympathetically and can describe their conditions to staff. | Can make sensible replies when spoken to | Can hold conversations and can take messages | Can follow and give simple descriptions and explanations | Can communicate effectively with a range of people in a variety of situations | Can present a logical and effective argument. Can analyse others' arguments |
| | READING | Can read technical books on nursing (for her project work). | Can read words and short phrases | Can read straightforward messages | Can follow straightforward written instructions and explanations | Can understand a variety of forms of written materials | Can select and judge written materials to support an argument |
| | WRITING | Writes out her logbook and project notes accurately and clearly. | Can write words and short phrases | Can write straightforward messages | Can write straightforward instructions and explanations | Can write reports describing work done | Can write a critical analysis, using a variety of sources |
| | USING SIGNS AND DIAGRAMS | Understands, with a little guidance, charts setting out emergency procedures. | Can recognise everyday signs and symbols | Can make use of simple drawings, maps, timetables | Can make use of basic graphs, charts, codes, technical drawings, with help | Can interpret and use basic graphs, charts, technical drawings unaided | Can construct graphs and extract information to support conclusions |
| | COMPUTER APPRECIATION | N/O | Can recognise everyday uses of computers<br>N/O | Can use keyboard to gain access to data | Can enter data into the system using existing programs | Can identify potential applications for computers | Can construct error free programs |
| PRACTICAL & NUMERICAL | SAFETY | Suggested moving an electric cable, so that it could not be tripped over. | Can explain the need for safety rules | Can remember safety instructions | Can spot safety hazards | Can apply safe working practices independently | Can maintain, and suggest improvements to, safety measures |
| | USING EQUIPMENT | Uses a mixer and knives for kitchen work (Supervised) | Can use equipment safely to perform simple tasks under guidance | Can use equipment safely to perform a sequence of tasks after demonstration | Can select and use suitable equipment and materials for the job, without help | Can set up and use equipment to produce work to standard | Can identify and remedy common faults in equipment |
| | NUMERACY | Scaled up a cake recipe by doubling the quantities | Can count objects | Can solve problems by adding and subtracting | Can solve problems by multiplying and dividing | Can calculate ratios, percentages and proportions | Can use algebraic formulae |
| SOCIAL | WORKING IN A GROUP | Always keeps other members of the nursing team informed about patients. | Can cooperate with others when asked | Can work with other members of the group to achieve common aims | Can understand own position and results of own actions within a group | Can be an active and decisive member of a group | Can adopt a variety of roles in a group |
| | ACCEPTING RESPONSIBILITY | Uses own initiative when cleaning out patients' rooms. | Can follow instructions for simple tasks and carry them out under guidance | Can follow instructions for simple tasks and carry them out independently | Can follow a series of instructions and carry them out independently | Can perform a variety of tasks effectively given minimal guidance | Can assume responsibility for delegated tasks and take initiative |
| | WORKING WITH CLIENTS | Can recognise if patients are uncomfortable and turn them in bed. | Can help someone to carry out clients' requests | Can carry out clients' requests under supervision | Can carry out clients' requests without supervision | Can anticipate and fulfil clients' needs from existing resources | Can suggest realistic improvements to services for clients |
| DECISION-MAKING | PLANNING | Can demonstrate how to turn a patient in bed. | Can identify the sequence of steps in everyday tasks, with prompting | Can describe the sequence of steps in a routine task, after demonstration | Can choose from given alternatives the best way of tackling a task | Can modify/extend given plans/routines to meet changed circumstances | Can create new plans/routines from scratch |
| | OBTAINING INFORMATION | Used the college library to find information for her project. | Can ask for needed information | Can find needed information, with guidance | Can use standard sources of information | Can extract and assemble information from several given sources | Can show initiative in seeking and gathering information from a wide variety of sources |
| | ASSESSING OWN RESULTS | Can recognise whether or not her work is satisfactory, both when nursing and cleaning kitchen ware. | Can receive advice about own performance | Can seek advice about own performance | Can assess own results, with guidance | Can assess own results for familiar tasks, without help | Can assess own performance and identify possible improvements |

N/O = No opportunity to assess

**Main Activities:**

1. Helping to look after geriatric patients.
2. Assisting in the nursing home kitchen.

153

Signed (Trainee/Student) Gillian Anderson

Signed (Supervisor/Tutor) J Quilley

Name of Centre and Course Ambleside Nursing Home / Pre-Nursing

Period covered by this profile (dates) — start: 10-9-83 end: 12-11-83

© 1988 City and Guilds of London Institute.

Mark III

**Final profile report**

The Final Profile issued when a student leaves the scheme—is comprised of four pages. The front page provides details of both student and scheme; the inside pages summarise the best examples of ability at the highest levels attained on the grid; and the back page describes how the profile system works and provides additional information such as other qualifications achieved.

The following 4 pages are an example of the Final Profile developed by the City and Guilds of London Institute.

Figure 5

# CITY AND GUILDS OF LONDON INSTITUTE

Incorporated by Royal Charter Founded in 1878

## Profile Report

NAME: GILLIAN ANDERSON AGE: 17

SCHEME: YOUTH TRAINING SCHEME
COMMUNITY SERVICE AGENCY
69 HIGH STREET
AMBLESIDE

FROM: AUGUST 1983 TO: MAY 1984

WORK EXPERIENCE:

COMMUNITY AND HEALTH SERVICE WORK

AMBLESIDE NURSING HOME
LAKEVIEW OLD PEOPLE'S HOME

DOMESTIC DUTIES - CLEANING : LAUNDRY : KITCHEN AND DINING ROOM
PERSONAL CARE OF RESIDENTS
OCCUPATIONAL THERAPY UNIT - HELPING RESIDENTS WITH KNITTING, SOFT TOY MAKING AND OTHER CRAFTS

OFF THE JOB LEARNING:

ENNERDALE COLLEGE OF FURTHER EDUCATION

INTRODUCTORY CARING COURSE -

(i) WORKING IN THE COMMUNITY
(ii) NUMERACY/LANGUAGE
(iii) NUTRITION AND HEALTH EDUCATION
(iv) COMMUNITY SERVICES

---

This is an end-of-scheme profile, prepared on the basis of regular consultations between scheme staff and the trainee. Profiles are completed in conformity with CGLI requirements, and this process (but not individual attainment levels) is monitored by CGLI.

## SPECIFIC APPLICATIONS OF ABILITIES

### COMMUNICATION

Listened sympathetically to patients worries and discussed their problems with them.

Wrote letters for short sighted patients.

Maintained laundry stock records and checked written orders.

Followed written instructions/patterns for knitting and soft toy making.

Prepared an illustrated booklet on 'Caring for the Elderly' after researching material from a variety of sources.

### SOCIAL ABILITIES

Worked well with other trainees and staff at Nursing Home - discussed patients progress.

Bathed, fed, dressed and shaved elderly patients without supervision.

Allocated and supervised work in occupational therapy unit.

Quickly got to know clients at each placement - reacted sensitively to the needs of confused elderly patients.

### PRACTICAL & NUMERICAL ABILITIES

Used kitchen, laundry and cleaning equipment; also variety of craft tools.

Used clinical thermometer; took blood pressure.

Handled wheelchairs correctly with regard to safety and comfort of occupant.

Cooked meals without supervision, weighing and measuring ingredients accurately.

Estimated and costed materials for craft projects.

### DECISION-MAKING ABILITIES

Planned recreational programmes to suit individual patients.

Helped organise patients outing - planning bus routes and researching required information on facilities.

Coped calmly with patients having epileptic fits.

**Figure 5** continued

## ATTAINMENTS IN BASIC ABILITIES

(Highest level from Progress tracking grid – see overleaf).

Bar indicates highest level attained

### COMMUNICATION

| | |
|---|---|
| TALKING AND LISTENING | Can communicate effectively with a range of people in a variety of situations |
| READING AND WRITING | Can use instruction manuals and can write reports describing work done |
| USING SIGNS AND DIAGRAMS | Can, after guidance, make use of basic graphs, charts, tables, drawings, etc. |

### SOCIAL ABILITIES

| | |
|---|---|
| WORKING IN A GROUP | Can be an active and decisive member of a group |
| WORKING WITH THOSE IN AUTHORITY | Can assume responsibility for delegated tasks and take initiatives |
| WORKING WITH CLIENTS | Can anticipate and respond to client's needs |

### PRACTICAL & NUMERICAL ABILITIES

| | |
|---|---|
| USING EQUIPMENT | Can select and use suitable equipment and materials for the job, without help |
| CONTROL OF MOVEMENT | Can use tools to do fine and/or complex work |
| MEASURING | Can measure out specified quantities of material by length, weight, etc. |
| CALCULATING | Can use x and ÷ to solve whole number problems |

### DECISION-MAKING ABILITIES

| | |
|---|---|
| PLANNING | Can modify/extend given plans/routines to meet changed circumstances |
| SEEKING INFORMATION | Can show initiative in seeking and gathering information from a wide variety of sources |
| COPING | Can cope with unexpected or unusual situations |
| ASSESSING RESULTS | Can assess own performance and identify possible improvements |

**Figure 5** continued

**HOW PROFILES ARE COMPILED:**

During the scheme the trainee has recorded his/her work and experience in a Log Book. From this record, specific examples of abilities in the four main areas (and the fourteen sub-categories) were identified and recorded at regular intervals against the **PROGRESS TRACKING GRID** shown below.

Simple tasks under supervision → Complex tasks without supervision

| | | | | | | (High Level) | ✓ |
|---|---|---|---|---|---|---|---|
| COMMUNICATION | TALKING AND LISTENING | Can make suitable responses when spoken to | Can hold conversations with workmates. Can use telephone to take messages | Can follow and give simple directions and explanations | Can communicate effectively with a range of people in a variety of situations | Can present a topic of and effective argument. Can interest other 3 arguments | |
| | READING AND WRITING | Can read short sentences | Can read and write messages | Can follow and give straight forward written instructions and explanations | Can use instruction manuals and can write reports describing work done | Can select and criticise written data and use it to produce own written work | |
| | USING SIGNS AND DIAGRAMS | Can recognise everyday signs and symbols | Can explain signs and symbols | Can, after guidance, make use of basic graphs, charts, tables, drawings, etc | Can interpret and use basic graphs, charts, tables and drawings unaided | Can construct graphs etc. and extract information to support arguments | |
| SOCIAL ABILITIES | WORKING IN A GROUP | Can cooperate with others when told | Can work with other members of group to achieve common aims | Can understand own position and results of own actions within a group | Can be an active and decisive member of a group | Can lead a group | |
| | WORKING WITH THOSE IN AUTHORITY | Can follow instructions for simple tasks and carry them out under guidance | Can follow instructions for a simple task and carry it out independently | Can follow a series of instructions and carry them out independently | Can carry out a series of tasks effectively given minimal guidance | Can assume responsibility for delegated tasks and take initiatives | |
| | WORKING WITH CLIENTS | Can help someone to carry out a client's request | Can carry out client's requests under supervision | Can carry out client's requests without supervision | Can recognise client's needs | Can anticipate and respond to client's needs | |
| PRACTICAL & NUMERICAL ABILITIES | USING EQUIPMENT | Can use equipment safely to perform simple tasks under guidance | Can use equipment safely to perform several step tasks after demonstration | Can select and use suitable equipment and materials for the job without help | Can set up and maintain equipment | Can identify and remedy common faults in equipment | |
| | CONTROL OF MOVEMENT | Can lift, carry and set down objects | Can use everyday tools | Can use tools to produce work to given tunes | Can use tools to do fine and/or complex work | Can perform tasks requiring a high degree of control | |
| | MEASURING | Can sort objects into sizes | Can read simple scales and dials | Can measure out specified quantities of material by length, weight etc | Can read precision instruments | Can set up and use precision instruments to make accurate measurements | |
| | CALCULATING | Can count | Can add and subtract whole numbers | Can use + and - to solve whole number problems | Can add and subtract decimals and simple fractions | Can multiply and divide decimals and simple fractions | |
| DECISION MAKING ABILITIES | PLANNING | Can describe the sequence of steps in everyday tasks, with prompting | After demonstration, can identify the sequence of steps in a routine task | Can choose from given alternatives the best way of tackling a task | Can modify/respond given plans/routines to meet changed circumstances | Can create new plans/routines from scratch | |
| | SEEKING INFORMATION | Can ask questions | Can find needed information when guided | Can use standard sources of information | Can assemble information from several sources | Can show initiative in seeking and gathering information from a wide variety of sources | |
| | COPING | Can cope with everyday activities with help | Can cope with everyday problems. Seeks help if needed | Can cope with changes in familiar routines | Can cope with unexpected or unusual situations | Can offer sensitive and effective help to the group | |
| | ASSESSING RESULTS | Can seek advice about own performance | Can assess own results with guidance | Can assess own results for familiar tasks without help | Can assess own performance and identify possible improvements | Can assess the group's performance and help to improve it | |

Enter ✓ in end column if high level is exceeded

**ADDITIONAL INFORMATION**
(NOT covered by CGLI Profile Scheme)

Is well motivated and willing to learn and is a conscientious and careful worker. Handled cash for making small purchases for patients.

CSE - English (2); Biology (3); Modern Studies (4)

CGLI - Numeracy Level I

CGLI - Caring Skills Certificate

Voluntary work - fundraising for charitable bodies
Committee member of Youth Club.

CGE. 00467. ET2



158 BEST COPY AVAILABLE

## REFERENCES


Further Education Unit. Profiles. September 1982.

Further Education Unit. Profiles in Action. October 1984.

Kash. M. M., & Borich, G. D. Teachers, in Walberg, H. J. (Ed.) Improving Educational Standards and Productivity, McCutchan, 1982.

Wilson, A. F. Employers and school-based assessment. Victorian Institute of Secondary Educaton, July 1982.

## EXERCISE VII (1)

With the help of the examples provided in this Chapter develop a Profile Report Form which is relevant to a subject that you teach.

(There is no suggested answer for this exercise—the form should be the subject of group discussion during a staff development exercise.)

# INDEX

achievement tests, 18
affective domain, 25-6, 39, 54-5, 96-7
affective skills, 42
aim, 24
alternatives, 65
analysis items, 27, 34-5
application items, 27, 30-3
assessment, 2
  continuous, 9
  in TAFE, 9
  objective, 2
  subjective, 2
assignments,
  field work, 57
  project, 57
aural assessment, 74

behavioural objectives, 40
  limitations of, 41
  in TAFE, 42
  usefulness of, 40
  *see* objectives

classification of questions, 45-7
class tests, 9
cloze items, 63
cognitive domain, 25-6, 54-5, 96-7
cognitive skills, 42
completion tests, 63
composite scores, 138-9
comprehension items, 27-9
  extrapolation, 27
  interpretation, 27, 29
  translation, 27, 29
conditions of assessment, 98
consensus moderation, 116
content validity, 127
continuous assessment, 10
criterion-referenced assessment, 10-12
  future of, 14
  grades, 13
  level of specificity, 15
  minimal acceptable performance, 15
  reliability, 13
  reporting results of, 139
  scores, 13
  statistics, 114-5
  validity, 13
criterion-referenced tests, 11, 19
  future of, 14

diagnostic test, 18
difficulty (of test items), 110
disclosed topic test, 75
discrimination index, 111
distracter analysis, 112
distracters, 65

educational objectives *see* objectives
error in reliability, 130
essays, 59, 78
  guidelines for preparing, 61-2
  in TAFE, 59
  open-ended, 60-1
evaluation, 7
  formative, 8
  qualitative, 8
  quantitative, 7
  research and, 8
  summative, 8
evaluation items, 27, 37-8
examinations, 9
  end of year, 9
  Year 12, 139

face validity, 127
facility (of test items), 110
formative evaluation, 8

grades, 13
goal, 24

item analysis, 109
  computer programs for, 114
  phi co-efficient, 118-121
item banks, 53
item difficulty, 110
item facility, 110
item review, 82-85

key, 65
knowledge items, 27-8

learning objectives, 101
listening assessment, 74-5

mastery, 55
matching pairs questions, 72
  guidelines for writing, 73
measurement, 3
modular approach, 98
multiple choice questions, 64-71, 78
  guidelines for writing, 66-71

norm-referenced assessment, 10-12
  test items, 12
  grades, 13
  scores, 13
  reliability, 13
  validity, 13
norm-referenced tests, 11, 19
  future of, 14
  statistics, 110
  *see* item analysis

objectives, 15, 101
  classification of, 25, 44-7
  educational, 23
  general, 24
  learning, 101
  levels of specificity, 24
  performance, 24-5
  record keeping, 55
  *see* behavioural objectives
objective test, 64
objectivity of tests, 51-2
open book test, 75
oral assessment, 74, 81

performance, 98
performance levels, 24
phi co-efficient, 118-121
practical assessment, 54-57
  checklists, 56
  self assessment, 57
profile reporting, 139-155
project and field work assignments, 57, 80
  choosing projects, 58
  group work, 58
  role of teachers, 58
  scope, 58
predictive validity, 128
psychomotor domain, 25-6, 40, 54-5
psychomotor skills, 42

qualitative evaluation, 8
quality of workmanship, 55
quantitative evaluation, 7

reliability, 125-6, 129
  co-efficient, 130
  error, 130
  of norm-referenced assessment, 13
  of criterion-referenced assessment, 13
reporting marks and grades, 138
reporting results, 133
  in TAFE, 133
report profiles, 139-155
reports, 133
  and admissions officers, 137
  and composite scores, 138-9
  and employers, 137
  and parents, 135
  and standards, 138
  and students, 134
  and teachers, 137
  audiences for, 134
research, 8

scores, 13
selection, 128
self assessment, 57
short answer questions, 63
  guidelines for preparing, 64
standards, 55, 98, 138
statistics (of tests), 109
  for criterion-referenced tests, 114
  for norm-referenced tests, 110
  storage, 113, 115
stem, 65
stimulus material, 65
structured essay, 62
  guidelines for preparing, 63
summative assessments, 9
summative evaluation, 8
synthesis items, 27, 36-7

table of specifications, 96-7, 100-101
test construction, 97
  use of time, 52-3
test planning, 94
  achieving balance, 94
test items, 51
test questions, 51
tests, 3, 51
  achievement, 3
  aptitude, 6
  criterion-referenced, 11, 19
  diagnostic, 7
  for class, 9
  norm-referenced, 11, 19
test specifications, 93-102
  assigning weightings, 96, 101
  comprehensiveness, 98
test types, 52-4, 62-71, 82-5
  listening assessment, 54, 74
  matching pairs, 72
  multiple choice, 52, 64-71, 78, 82-5
  open-ended essay, 52, 60
  oral assessment, 54, 74
  practical assessment, 54
  project work, 54
  sentence completion, 52, 64
  short answer, 52, 63
  structured essay, 52, 62
  true-false, 71
  written assessment, 54
true-false questions, 71
  guidelines for writing, 71-2

validity, 109, 125-6
  and TAFE teacher, 125
  content, 127
  face, 127
  predictive, 128
  via consensus moderation, 116

written assessments, 59
  essay, 59
  matching pairs, 59
  multiple choice, 59
  short answer, 59
  structured essay, 59
  true-false, 59

Year 12 examination, 129

STUDENT ASSESSMENT: A HANDBOOK FOR TAFE TEACHERS

This book is for teachers and trainers working in vocational education.

It aims to meet the needs of TAFE teachers as well as trainers from industry and commerce who wish to improve their assessment techniques.

This is a practical book with copious examples of what to do and what not to do when testing students.

It should be equally of value to beginners and to those who wish to sharpen their existing skills in assessment.

Readers with a knowledge of other texts on the subject will note that new ideas and new emphases are presented. Practical assessment receives special attention as do new approaches to describing student performance.

Peter Thomson has worked in the area of student assessment for the past 15 years both in Australia and overseas. He was formerly Head of the Test Development Division at the Australian Council of Educational Research (ACER) and is currently the Deputy Director of the TAFE National Centre for Research and Development.

NELSON WADSWORTH

ISBN 0 17 006896 X